





1151 TGATGCCCCCTGGATATCTGCCACCCCTGCTTTGGCGGATTAATCTGC 1200  
 |||||||  
 384 almetProleuspileleuProthleuLeuAArgAspleuileval 400  
 |||||||  
 1201 GGCGATTACCGACAGCGCGATGGTGTGCTTGAATTGGACGAGA 1230  
 |||||||  
 401 GYAAPTThrAspSerAlaGlnAlaLeuGlyCysLeuGluLeuAspGluGI 417  
 |||||||  
 1251 AGACCTGCGTTGTGACGCTTGTGCGCGGCAAAATACGAATAGAGGCC 1300  
 |||||||  
 417 uaspleuAlaLeuGlyCysSerPheValCysProGlyLysTyrGluTyrGI 434  
 |||||||  
 1301 CGCTGTGGCGCAAGTGTGGAACCATGAGAGAGAGC 1341  
 |||||||  
 434 roleuLeuArgLysValLeuGluThrIleGluLysGluGI 447  
 |||||||

seq\_name: Swissprot\_40: NORA\_NEIMA

seq\_documentation\_block:

ID NORA\_NEIMA STANDARD; PRT; 447 AA.  
 AC Q9JVP8;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)  
 DE (Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A) (NQR complex subunit A) (NQR-1 subunit A).  
 GN NORA OR NMA0752.  
 OS Neisseria meningitidis (serogroup A).  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID:55699;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-22491 / SEROGROUP A / SEROTYPE 4A;  
 RX MEDLINE-20222556; PubMed-10761919;  
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C., Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T., Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S., Jørgensen K., Leather S., Moule S., Mungall K., Quail M.A., Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J., Whitehead S., Spratt B.G., Barrall B.G.;  
 RA "Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491."  
 RL Nature 404:502-506(2000).  
 CC -1- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO NQRE ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF UBISEMIQUINONE TO UBIQUINOL (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+) (IN) = NAD(+) + UBIQUINOL + NA(+) (OUT).  
 CC -1- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NQRB, NQRC, NQRD, NQRE AND NQRF (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE NORA FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AL162754; CAB84035.1;  
 KW Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport; Complete proteome.  
 SQ SEQUENCE 447 AA; 48673 MW; B808CAA82C901A1D CRC64;

alignment\_scores:

Quality: 2251.00 Length: 447  
 Ratio: 5.036 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 97.763

alignment\_block:  
 US-09-303-518d-125 x NORA\_NEIMA ..

Align seq 1/1 to: NORA\_NEIMA from: 1 to: 447

1 ATGATTAAATCAAAAAAGGTCTAAACCTGCCATCGCGGCGAGACCGA 50  
 |||||||  
 1 MetIleLysIleLysLysGlyLeuAsnLeuProIleAlaGlyArgProI 17  
 |||||||  
 51 GCAGCCGTTTACGAGCGCCGCCCATTTACGGAATGCGGTGCTGGCG 100  
 |||||||  
 17 uGlnValIleTyrAspGlyProValIleThrGluValAlaLeuLeuGI 34  
 |||||||  
 101 AAGAAATGCGCGGTATGCGCCCTCGATGAATCAAGAGAGCGATGGC 150  
 |||||||  
 34 IuGluTyrAlaGlyMetArgProSerMetLysValLysGluGlyAspAla 50  
 |||||||  
 151 GTCAAAAAGGCCAAGTGTCTTGAAGACAAAAAAGATCCGGCGTGT 200  
 |||||||  
 51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyVal 67  
 |||||||  
 201 GTTTACTGCGCGGCTTCAAGCAAAATCGCGCATTCACCGTGCGAAA 250  
 |||||||  
 67 IPhetThrAlaProValSerGlyLysIleAlaIleHisArgGlyGlu 84  
 |||||||  
 251 AGCGCGTACTTCAGTCAGTCGTGATTCGCTGAGGCAACGACGAATC 300  
 |||||||  
 84 YsaArgValLeuGlnSerValIleLeuAlaValGluLysAsnAspGluIle 100  
 |||||||  
 301 GAGTTTGAACGCTACGACCTGAAGCGCTGCAAACTTAAGCGCGAGAGA 350  
 |||||||  
 101 GluPheGluArgTyrIleArgProGluAlaLeuAlaAsnLeuSerGlyGlu 117  
 |||||||  
 351 AGTGGCGCGCAACCTGATCCATCCGCTTGTGACTGCGTGCACACC 400  
 |||||||  
 117 uValArgArgAsnLeuIleGlnSerGlyLeuThrAlaLeuArgThr 134  
 |||||||  
 401 GTCCGTTCAGCAAAATTCCTGCGCTGATGCCAGCGCTTGGCATCTTC 450  
 |||||||  
 134 rGProPheSerLysIleProAlaValAspAlaGluProPheAlaIlePhe 150  
 |||||||  
 451 GTCAATCGATGAGACCAATCCGTGCGCGCGACCTAAGCGCATATT 500  
 |||||||  
 151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProValValIle 167  
 |||||||  
 501 CAAAGAGCCCGCGAGATTTCAAACGCGGCTGTGATAGGCGGTT 550  
 |||||||  
 167 eLysGluAlaAlaGluAspPheArgGlyLeuLeuValLeuSerArgL 184  
 |||||||  
 551 TGACCGAAGCAAAATCCATGTTTGAAGCAACCTGGCGCAGACGTGG 600  
 |||||||  
 184 eThrGluArgLysIleHisValCysLysAlaIleGlyAlaAspValPro 200  
 |||||||  
 601 TCTGAAATGCTGCCAATCGAAACATGATGATTTGGGCGCGCATCC 650  
 |||||||  
 201 SerGluAsnAlaAlaAsnIleGluThrHisGluPheGlyProHisr 217  
 |||||||  
 651 TGCCGGTTGATGAGCAGCAGCATTCATTCATGAGCGGTGGCGGCA 700  
 |||||||  
 217 oAlaGlyLeuSerGlyThrHisIleHisPheIleGluProValGlyAla 234  
 |||||||  
 701 ATAAACCGGTGACCATCAATTAACAGATCTAATACATTTGGCGCT 750  
 |||||||  
 234 snLysThrValIleThrIleAsnTyrGlnAspAlaIleAlaIleGlyArg 250  
 |||||||  
 751 TTGTTGCAACAGCGCGTCTGAACACGAGCGGTGATTCCTAGAGTGG 800  
 |||||||  
 251 LeuPheAlaThrGlyArgLeuAsnThrGluArgValIleAlaLeuGly 267  
 |||||||  
 801 TTCTCAAGTCACAAACCGCGCTCTGGCGTACCGTTTGGGTGGCAAG 850  
 |||||||  
 267 ySerGlnValAsnLysProArgLeuLeuArgThrValLeuGlyAlaLysV 284  
 |||||||





## alignment\_scores:

Quality: 1629.00 Length: 448  
 Ratio: 4.093 Gaps: 2  
 Percent Similarity: 88.839 Percent Identity: 68.973

## alignment\_block:

US-09-303-518D-125 x NORA\_HAEIN ..

Align seg 1/1 to: NORA\_HAEIN from: 1 to: 447

```

1 ATGATTAAATCAAAAAGATCTAAACCTGCCATCGCGGCGAGACCGGA 50
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 MetIleThrIleLysLysGlyLeuAspLeuProIleAlaGlyLysProI 17
51 GCAAGCCGTTTACGAGCGCCGCCATTACCGAAGTCGCGTGTGGCG 100
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 aglnValIleHisSerGlyAsnAlaValasnGlnValAlaIleLeuGly 34
101 AAGAATATGCCGATATGCCGCCCTCGATGAAGTCAAGAGCGCATGCC 150
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 lueGlyValGlyMetArgProSerMetLysValArgGlyLysPval 50
151 GTCAAAAAGGCCCAAGTCTGTTTGAAGACAAAAGAAATCCGGGGTGT 200
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 ValLysLysGlyGlnValIleuPheGlyAspLysLysAsnProGlyVal 67
201 GTTTACTGCGCGCGCTTACGAGCAAAATCGCGCGATTACCGTGGCGAA 250
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
67 ePheThrAlaProIleSerGlyThrIleThrAlaIleAsnArgGlyLul 84
251 AGCGGCTACTGATGATGATGATGATGATGATGATGATGATGATGATG 300
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
84 ysaArgValLeuGlnSerValValIleAsnValGlnGlyAspGlyLys 100
301 GAGTTTGAACCTACGACCACTGACCGTGGCAAACTTAAGCGCGGAGA 350
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 ThrPheAlaLysTyrSerThrGlnLeuAsnThrLeuSerSerGlnG 117
351 ACTGCGCCGCAACCTGATCAATCGGTTTGTGACTGCGTGGCGACCC 400
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
117 nValLysGlnAsnLeuIleGlnSerGlyLeuThrPheAlaLeuArgThr 134
401 GTCCGTTACGCAAAATCTCGCCGTCGATCGCGACCGCTGGCCATCTTC 450
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134 rGProPheSerLysValProSerIleGlnSerGlnAlaSerSerLePhe 150
451 GTCAATGCGATGACACCAATCGCTGGTGGCGCGACCGTATAT 500
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProSerValVal 167
501 CAAAGAGCCCGCGAGATTTCAACGCGCGCTGTGTGATTTGAGCCGTT 550
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
167 ulysGlyTyrSerGlnAspPheThrAsnGlyLeuThrValLeuSerArg 184
551 TACCGAAGCGCAA...ATCCATGTTTGTAAAGCAGCTGGCGGACAGCTG 597
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
184 eurPheProSerLysProLeuHisLeuGlyLysAlaGlyAspSerAsnIle 200
598 CCGTGTGAATAATGCTCCCAACATCGAACACATGATGATGCGGCGCCGA 647
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
201 ProThrAlaAspLeuGlnAsnLeuGlnIleHisAspPheThrGlyVal 217
648 TCCGTCGCGTTTGAAGTGGCAGCAGCATTTATTCATCGAGCGGTCGCG 697
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
217 sProAlaGlyLeuValGlnGlyThrHisIleHisPheIleAspProVal 234
698 CGAATAAACCGCTGTGACATCAATTTCAAGATGTAATTAATCAATGGC 747
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
234 leglnLysThrValThrHisIleHisnLysnLysnLysnLysnLysn 250
748 CGTTTGTTCAGACAGCGCGTGTGTAACGAGCGCGGCGTGGCTAGG 797
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
251 LysLeuPheThrThrGlyLeuLeuTyrSerGlnArgValIleSerLeuAl 267

```

```

798 TGGTTCTCAAGTCAACAACCGCGCTCTTGCGTACCGCTTTGGGTGCGA 847
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
267 aglyProGlnValLysGlnProArgLeuValArgThrThrIleGlyAla 284
848 AAGTATCGCAAAATTAATGCGCGCGCAATTTGTTGACACAGCAACCGCGT 897
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
284 snLeuSerGlnLeuThrGlnAsnGlnLeuSerAlaGlyLysAsnArgVal 300
898 ATTTCGCGTTCGATTTGAACGCGCGCATTTACACAGCGCGCGCATTTA 947
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
301 lIleSerGlySerValLeuGlyGlnIleAlaLysAspSerHisAspTy 317
948 TTTGGGACGCTTACCAATCAGATTTCCGTTATCGAAGAGCGCGGACGA 997
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
317 rLeuGlyArgGlyThrAlaLeuGlnValSerValIleAlaGlnGlyAsn 334
998 AAGAGCTGTTCGCGTGGTTCGCGCGAGCGGCAAAATCTCCATCAGC 1047
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
334 ysgLupPhePheGlyTTrpIleMetProGlnAlaAsnLysTyrSerVal 350
1048 CGTCAACCGCTCGCGCATTTCCGTAACAAACACTCTTCAAGTTCAAC 1097
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
351 ArgThrValLeuGlnLysPheSerLys...LysLeuPheAsnPheThr 366
1098 AGCGGTCACGCGCGCGACCGCGCGCATGTCGCGATTTGATTAAGAC 1147
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
366 rSerGlnAsnGlyGlyLysArgLysMetValProIleGlySerTyrGln 383
1148 GCGTGAATGCCCTTGGAATCTTGCACCGCTGCTTTGGCGCATTAATC 1197
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
383 rValMetProLeuAspIleLeuProThrLeuLeuArgAspLeuIle 399
1198 GTGCGGATACCGACAGCGCGAGCATTTGGTTCGTTGGAATTTGACGA 1247
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
400 ValGlyAspThrAspGlyAlaGlnGlnLeuGlyCysLeuGlnLeuAsp 416
1248 AGAAGACCTCGCTTTGTCAGCTTTCGTCGCGCGCGCAATACGATTA 1297
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
416 uGlnAspLeuAlaLeuGlySerPheValLysProGlyLysTyrGlyTyr 433
1298 GCCCGCTGTTGCGCAAGTGTGTAACCAATGAGAGAGAGCGC 1341
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
433 LysIleLeuArgGlnValLeuAspLysIleGlnLysGlnGly 447
seq_name: swissprot_40:NORA_VIBHA
seq_documentation_block:
ID NORA_VIBHA STANDARD: PRT: 446 AA.
AC 09FWM1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
DE (Na(+)-translocating NOR subunit A) (Na(+)-NOR subunit A) (NOR complex
DE subunit A) (NOR-1 subunit A).
DE NORA.
OS Vibrio harveyi.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=669;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN-BB120;
RC MEDLINE-20056044; PubMed-10587447;
RA Zhou W., Bertsova Y.V., Feng B., Tsatsos P., Verkhovskaya M.L.,
RA Genis R.B., Bogachev A.V., Barquera B.;
RT "Sequencing and preliminary characterization of the Na+-translocating
RT NADH:ubiquinone oxidoreductase from Vibrio harveyi.";
RL Biochemistry 38:16246-16252(1999).
CC -!- FUNCTION: NOR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO MORE
CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF

```









```

601 TCTGAATGTCGACATCGAACAACATGATTCGGCGCCGCGATCC 650
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
200 ArgSerSerGlnSerAsnValGluGlnHisValPheAspGlyProHisP 216
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
651 TCCGCGTTGAGTGGCAGCAGCATTCATTCATCGAGCCGGTGGCGGA 700
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
216 CalaglyLeuAlaGlyThrHisMetHisPheLeuTyrProValAsnAlaG 233
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
701 ATAAACCGGTGGACCATTCATTCATCAAGTATTCATTCATTCGCGCT 750
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
233 LuAsnValAlaLeuTyrSerIleAsnTyrGlnAspValIleAlaPheGly 249
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
751 TTGTTGCAACAGCGCGCTGTGAACACGCGCGGATTCGCTTACGGGG 800
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
250 LeuPheLeuThrGlyGluLeuTyrThrAspArgValValSerLeuAlaG 266
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
801 TTCTCACTCAACAACACGCGCGCTTCGATTCGATTCGATTCGATTC 850
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
266 yProValValAsnProArgLeuValArgThrValIleGlyAlaSerL 283
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
851 TATCGCAATTCATTCGCGCGCATTCGATTCGATTCGATTCGATTC 900
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
283 euAspAspLeuThrAspAsnGluLeuMetProGlyGluValArgValIle 299
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
901 TCCGCGTTGATTCGACGCGCGCATTCACACAGCGCGCGATTCATTC 950
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
300 SerGlySerValLeuThrGlyThrHisAlaThrGlyProHisAlaTyrLe 316
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
951 GGGACGCTACACACATTCGATTCGATTCGATTCGATTCGATTCGATTC 1000
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
316 uGlyArgTyrHisGlnGlnValSerValLeuArgGluGlyArgGlyGly 333
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1001 AGCTGTGCGCGTGGTGGCGCGCGACGCGCAATTCATTCATTCATTC 1050
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
333 LuLeuPheGlyTyrPalaMetProGlyLysAsnLysPheSerValThrArg 349
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1051 ACAACCCCTGGGCATTCCTGAAAAACAACTTCATTCATTCATTCATTC 1100
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
350 SerPheLeuGlyHisValPheGlyGlyLeuPheAsnMetThrThr 366
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1101 CGTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1150
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
366 rThrAsnGlySerAspArgSerMetValProIleGlyAsnTyrGlyArg 383
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1151 TGATGCCCTTGATATCTCTGCCACCCGCTTCGCGCGATTCATTCATTC 1200
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
383 alMetProLeuAspMetGluProThrLeuLeuLeuArgAspLeuGlyAla 399
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1201 GGGCATACGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1250
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
400 GlyAspThrAspSerAlaGlnAlaLeuGlyAlaLeuGluLeuAspGlyG 416
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1251 AGACCTGCGTTGTCAGCTTCGTCGCGCGCGCGCGCGCGCGCGCGCG 1300
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
416 uAspLeuAlaLeuGlyThrPheValGlySerProGlyLysTyrGlyArg 433
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1301 CGGTGGCGCGCAACTGCTGGAACCATTCGAGAGGAGGCG 1341
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
433 hrLeuLeuArgGlyGlyLeuAspThrIleGlyLysGlyGly 446
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
seq_name: SwissProt_40: NORA_VIBCH

```

```

seq_documentation_block:
ID NORA_VIBCH STANDARD: PRT: 446 AA.
AC 09KPS1; 09X4Q3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
DE (Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A) (NQR complex
DE subunit A) (NQR-I subunit A).
GN NORA OR VC2295.
OS Vibrrio cholerae.

```

```

OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxId=666;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=EL TOR N16961 / SEROTYPE O1;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Yamashiro J., Bass S., Qin H., Dragol I., Sellers P.,
RA McDonald L., Uitterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RA cholerae."
RL Nature 406:477-483(2000).
CC -1- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO MORE
CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC UBISEMIOQUINONE TO UBIQUINOL.
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+) (IN) = NAD(+) +
CC UBIQUINOL + NA(+) (OUT).
CC -1- SUBUNIT: COMPOSED OF SIX SUBUNITS; NORA, NORB, NORC, NORD, NORE
CC AND NORE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NORA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF117331; AAD29962.1;
DR EMBL: AE004300; AAF95439.1; ALT_INIT.
DR TIGR: VC2295;
KW Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
KW Complete proteome.
SQ SEQUENCE 446 AA; 48624 MW; 428E8C397EB8A163D CRC64;

```

## alignment\_scores:

```

Quality: 1455.50 Length: 447
Ratio: 3.810 Gaps: 1
Percent Similarity: 85.459 Percent Identity: 61.074

```

## alignment\_block:

```

US-09-303-518D-125 x NORA_VIBCH ..

```

```

Align seg 1/1 to: NORA_VIBCH from: 1 to: 446

```

```

1 ATGATTAAATCAAAAAGGCTTAACCTGACCCATCGCGGCGACACCGGA 50
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 MetIleThrIleLysGlyLeuAspLeuProIleAlaGlyThrProSe 17
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 GCAGCGCGTTTACGAGCGCGCGCGCATTCACGAGTCCGCTTGGCG 100
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 rGlnValIleSerAspGlyLysAlaIleLysLysValAlaLeuLeuGlyG 34
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 AAGAAATATGCGGTATGCGCGCGCGCGCATGAAAGTGAAGGAGCGATGCC 150
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 LuGlyTyrValGlyMetArgProThrMetHisValArgValGlyAspGlu 50

```

```

151 GTCAAAAAGGCCAAGTGTGTTGAGACAAAAGAAATCCGGCGGTGT 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
51 ValLysLysAlaGlnIleLeuPheGlnLysLysAsnProGlyVal 67
201 GTTTACTGCGCGGCTTCAGGCAAAATCGCGCATTCACCGTGGCGAA 250
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
67 spherSerProValSerGlyLysValAlaGlnLeuAsnArgLysAla 84
251 AGCGGCTACTTCAGTCACTGCTGATGGCGTTGAAGCAGCAGCAATC 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
84 ysArgValLeuGlnSerValValIleGlnValAlaGlyAspSerIle 100
301 GAGTTGAACGCTTACGACCTGAGCGGTGCGCAAACTTAAGCGCGAGA 350
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
101 ThrPheAspLysPheGlnAlaAsnGlnLeuAlaSerLeuAsnArgAsp 117
351 AGTGGCGCGCAACCTGATCCAAATCCGTTGTGACCTGCGTGGCGACC 400
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
117 alleLysThrGlnLeuValAlaGlnSerGlyLeuThrPheAlaPheArg 134
401 GTCCGTTGAGCAAAATTCCTGCGCGTGCAGTGCAGCGCTTGCCATTC 450
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
134 rGProPheSerLysValProAlaIleAspSerThrSerGlnAlaIlePhe 150
451 GTCAATGCGAGTGACACCAATCCGCTGCGTGGCGACCTTACGTCATTAT 500
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
151 ValThrAlaMetAspThrAsnProLeuAlaGlnProThrValIle 167
501 CAAGACAGCGCGCGAGATTTCACACGCGCGCTGTGTTGATGAGCGCTT 550
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
167 eAsnGlnGlnSerGlnAlaPheValAlaGlyLeuAspPheLeuSerAla 184
551 TGACGCAAGCGCAAAATCCATGTTTGAAGCAGCTGCGCGCAGACGTGCG 600
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
184 euThrThrGlyLysValLysLys...LysGlyThrSerLeuPro 199
601 TGTGAAATGCTGCCACATGCAACACATGATTCGGCGCGCGCATCC 650
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
200 ArgSerGlnGlnProAsnValAlaGlnIleHisValPheAspGlyProHis 216
651 TGCCGTTGAGTGGCAGCAGCATTCATTCATGAGCGCGGTGCGCGCGA 700
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
216 cAlaLeuAlaGlnLysThrHisMetHisPheLeuThrProValSerAla 233
701 ATAAACCGTGTGACCATTAATTAAGATGTAATTAATTAATTAATTAAT 750
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
233 sPheIleValAlaThrSerIleAsnLysArgLysAspValIleAlaValGly 249
751 TTGTTTGAACAGCGCGCTGTAACACGCGCGGTGATGCGCTTACGTGG 800
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
250 LeuPheLeuThrGlyLysLeuThrGlnAlaValAlaSerLeuAlaGln 266
801 TTCTAGTCAACAAACCGCGCTTGTGCTGACGTTTGGTGGCGAAG 850
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
266 lProValValAsnLysProArgLeuValArgThrValMetGlyAlaSer 283
851 TATGCCAATTAATGCGCGCGAATGTTGACACAGACACCGCGTGATT 900
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
283 euGlnGlnLeuValAspSerGlnIleMetProGlyGlnValArgIle 299
901 TCCGGTGGTATGACGCGCGCATTCACAGCGCGCGCGCATTAATT 950
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
300 SerGlySerValLeuSerGlyThrLysAlaThrGlyProHisAlaTyr 316
951 GGGACGCTACACATCAGATTTCCTATGAGAGAGCGCGCGCAAG 1000
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
316 uGlyArgThrHisLeuGlnValSerValLeuArgGlnLysArgAsp 333
1001 AGCTTTGCGGTGGTGGCGCGCGCGCAACATTAATTAATTAATTAAT 1050
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
333 lLeuPheGlyTyrPheAlaMetProGlyLysAsnLysPheSerValThr 349
1051 ACAACCCTCGGCGCATTCCTGAAAAACAACACTTCAAGTTCAACACAGC 1100

```

```

350 SerPheLeuGlnLysIleLeuPheLysGlyGlnValTyrAsnMetThrThr 366
1101 CGTCACGCGCGCGCGCGCGCGCGCATGTCGCGCATGTTGTTACAGCGG 1150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
366 rThrAsnGlySerAspArgSerMetValProIleGlyAsnTyrGlyLys 383
1151 TGATGCCCTTGATATCTGCGCCACCGCTGTTTGGCGCATTAATCTGC 1200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
383 alMetProLeuAspMetGlnProThrLeuLeuLysArgLysPheGln 399
400 GlyAspSerAspSerAlaValArgLeuGlyAlaLeuGlnLeuAsp 416
1201 GCGCATACCGCAGCGCGCGCGCATGTTGTTGTTGTTGTTGTTGTTG 1250
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
400 GlyAspSerAspSerAlaValArgLeuGlyAlaLeuGlnLeuAsp 416
1251 AGACCTGCGTTTGGCAGCTGCTGCTGCGCGCGCAATGCAATGCGCC 1300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
416 uAspLeuAlaLeuLysThrPheValLysProGlyLysTyrGlyLys 433
1301 CGCTGTTGCGCAAGTGTGGAACCATTAAGAGAGAGAGAGAGAGAG 1341
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
433 lLeuLeuArgGlnLysLysLeuAspLysIleGlnLysGlnLys 446

seq_name: SwissProt_40:NORA_PSEAE
seq_documentation_block:
ID NORA_PSEAE STANDARD: PRT; 445 AA.
AC 09HK6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
DE (Na(+))-translocating NOR subunit A (Na(+)-NOR subunit A) (NOR complex
DE subunit A) (NOR-1 subunit A).
GN NORA OR PA2999.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAOI;
RX MEDLINE=20437337; Pubmed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltz L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reiter J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an
RT opportunistic pathogen."
RT Nature 406:959-964(2000).
RL -1- FUNCTION: NOR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO NQRE
CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC UBISEMIQUINONE TO UBIQUINOL (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+) (IN) = NAD(+) +
CC UBIQUINOL + NA(+) (OUT).
CC -1- SUBUNIT: COMPOSED OF SIX SUBUNITS: NQRA, NQRB, NQRC, NQRD, NQRE
CC AND NORF (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NORA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AE004724; AAC06387.1; -
KW Oxidoreductase; NAD; ubiquinone; Transport; Sodium transport;

```





transport, social transport;

1. 2. 3. 4. 5. 6. 7. 8. 9. 10. 11. 12. 13. 14. 15. 16. 17. 18. 19. 20. 21. 22. 23. 24. 25. 26. 27. 28. 29. 30. 31. 32. 33. 34. 35. 36. 37. 38. 39. 40. 41. 42. 43. 44. 45. 46. 47. 48. 49. 50. 51. 52. 53. 54. 55. 56. 57. 58. 59. 60. 61. 62. 63. 64.

229 LeHisIleHisSerValAlaProIleThrAsnGluGluValAlaPhe 245

```

715 ACCATCAATTATCAAGATGTATACATGTCGCGCTTTGTTGCAACAGG 764
|||||.....|
246 ThrLeuSerPheGlnAspValLeuThrIleGlyHisLeuPheLeuYsgI 262
|||||.....|
765 CCGCTGTCAACACGACGCGCGATGATGCGCTTCAAGTCACACA 814
|||||.....|
262 YargIleuHisIsgIuInValIThrAlaLeuAlaGlyThrAlaLeuYsg 279
|||||.....|
815 AACCG.....CGCCTCTTGGCAGCTTTGGGTGGCAAGATATGCGAA 858
|||||.....|
279 erSerLeuArgArgIYrValIleThrThrIlysgIYalaserPheSer 295
|||||.....|
859 ..ATTATGCGGGGGAATGGTGTGACACAGCAACCGCGATTTCCGG 905
|||||.....|
296 LeuIleAsnLeuAsnAspIleSerAspAsnAspThr...LeuIleSerG 311
|||||.....|
906 TTCGGTATTGAACGCGCGCATTTACACAGCGCGCAGCAT...TATTTGG 952
|||||.....|
311 YAspProLeuThrGlyArgLeuCysIlysgIuGluProPheLeuG 328
|||||.....|
953 GACGTTACCAATCAGATTCCTGTTATCGAAGAAGCGCGACGCAAGAG 1002
|||||.....|
328 LypHeArgAspHisSerIleSerValLeuHisAsnProThrLysArgGlu 344
|||||.....|
1003 CTGTTGCGGTGGGTGGCGCGCGACCGCAATATCTCATCAGCGCTAC 1052
|||||.....|
345 LeuPheSerPheLeuArgIleGlyLysAsnLysProThrPheThrIly 361
|||||.....|
1053 AACCTCGCGCATTTCTGAAACAACTCTCAAGTTCAAC.....A 1096
|||||.....|
361 rTyIleuSerGlyPhePheLysLysArgThrTyThrAsnProAsp 378
|||||.....|
1097 CAGCGCTCAAGCGCGCGCGCGCGCGCATGTCGATTCATTCAGAG 1146
|||||.....|
378 hAsnLeuHisIleYgluThrArgProIleIleAspThrAspIleTyAsp 394
|||||.....|
1147 CGCGTGATGCGCTTGGATATCTCGCGCACCGCTTTGGCGCATTAAT 1196
|||||.....|
395 LysValMetProMetArgIleProValValProLeuIleLysAlaValI 411
|||||.....|
1197 CGTCGCGCATTCACGACGCGCGCGCATTTGGGTGCTTGAATTCGACG 1246
|||||.....|
411 eThrLysAsnPheAspLeuAlaAsnGluLeuGlyPheLeuGluValCysG 428
|||||.....|
1247 AAGAGACCTGCTTGTGCAAGCTTCGTCGCGCGCGCAATACGAATAC 1296
|||||.....|
428 LysGluAspPheAlaLeuProThrIleAspProSerLysThrGluMet 444
|||||.....|
1297 GGCCCGCTGTGCGCAAGTGTGGAACCATTCGAGAAGAA 1338
|||||.....|
445 LeuThrIleValLysGluSerLeuIleGluIlyrAlaLysGlu 458
|||||.....|
seq_name: SwissProt_40:NORA_CHLTR

```

```

seq_documentation_block:
ID_NORA_CHLTR STANDARD; PRT; 465 AA.
AC 084639;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Probable Na(+)-translocating NADH-quinone reductase subunit A
DE (EC 1.6.5.-) (Na(+)-translocating NOR subunit A) (Na(+)-NOR subunit A)
DE (NOR complex subunit A) (NOR-1 subunit A).
GN NORA OR NOR1 OR CT634.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/DW-3/CX;
RX MEDLINE=99000809; PubMed=9784136.
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,

```

```

RA Mitchell W.P., Olinger L., Ratusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.:
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
CC -1- FUNCTION: NOR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO NORE
CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC UBISEMIOQUINONE TO UBIQUINOL (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+)(IN) = NAD(+) +
CC UBIQUINOL + NA(+)(OUT).
CC -1- SUBUNIT: COMPOSED OF SIX SUBUNITS; NORA, NORE, NORE, NORE
CC AND NORE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NORA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC
CC -----
CC EMBL: AE001334; AAC68238.1; -.
CC Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
CC Complete proteome.
CC KW
CC SEQUENCE 465 AA; 51757 MW; 303F88DC133AB998 CRC64;

```

```

alignment_scores:
Quality: 446.00 Length: 464
Ratio: 1.517 Gaps: 14
Percent Similarity: 63.362 Percent Identity: 28.233

```

```

alignment_block:
US-09-303-518d-125 x NORA_CHLTR ..

```

```

Align seg 1/1 to: NORA_CHLTR from: 1 to: 465

```

```

4 ATTAATCAATAAAGGCTTAACCTGTCATCGCGCGACGCGAGACA 53
|||||.....|
3 IleValValSerArgIlyleuAspLeuSerLeuYsgIYalProLysI 19
|||||.....|
54 AGCGCTTTACGACGCG.....CGGCGCATTCACGAATGCGCGTTC 94
|||||.....|
19 uSerGlyPheCysGlyLysValAspProThrIlyrValSerValAspLeu 35
|||||.....|
95 TTGGCGAAGATATGCGCGATGCGCGCGCGCATGATGAAGTCAAGGAAGC 144
|||||.....|
36 ....ArgProPheAlaProLeuProLeuGlyValIlyValThrProGlu 50
|||||.....|
145 GATGCGCTCAATAAAGGCGCAAGTGTGTTGAAGACAAAGAAATTCGCG 194
|||||.....|
51 AspGluValThrAlaGlySerProLeuAlaGluIlyrLysLeuPheSerI 67
|||||.....|
67 YValPheIleThrSerProValAspGlyGluValValGluIleArgG 84
|||||.....|
245 GCGAAGAGCGCGCTTCACTGATGCTGATGCTGATGCTGATGCTGATG 291
|||||.....|
84 LysIlyAspGluAlaLeuLeuGluIleValIleLysLysLysProGlyIle 100
|||||.....|
292 GACGAATCGATTTGATGACGCTGACGACCTGGAAGCGCTGCAACTTAG 341
|||||.....|
101 SerGluThrIlyPheSer.....TyAspLeuGlnSerLeuTh 113
|||||.....|
342 CCGCGAAGAGTGGCGCGCACTGATTCATTCGATTCGATTCGATTCGCG 391
|||||.....|
113 rGluLysAspLeuLeuGluIleValPheLysLysGluIlyLeuPheAlaLeuP 130
|||||.....|
392 TGCGACACCGCTCGCTTCAGCAAAATTCCTGCGCTGATGCGCGAG...CCG 438

```

```

130 heylsGlnArgProPheAsp...IleProAlaLeuProThrGlnSerPro 145
      ::::: |||||::: |||||::: ::::: ||
439 TTGGCCATTTTCGTAATGGCAGACCAACCAATCCGCTG...GCTGC 482
      ::::: |||||::: |||||::: ::::: ||
146 ArgAspValPheIleAsnIleAsnIleAsnArgProPheThrProSerIa 162
      ::::: |||||::: |||||::: ::::: ||
483 CGACCCATGAGGTGATTAACAAGACCCGCGAGAT.....T 520
      ::::: |||||::: |||||::: ::::: ||
162 IGILySHISLeuSerLeuPheSerLeuSerIySHISpGILyTYrILeP 179
      ::::: |||||::: |||||::: ::::: ||
521 TCAAGCGCGCGCTGTGTATGAGCGGCTTGACGACGCAACCAATTCAT 570
      ::::: |||||::: |||||::: ::::: ||
179 heValValGILyValGILnAlaIleAlaLysLeuPheGILyLeuLysProHIS 195
      ::::: |||||::: |||||::: ::::: ||
571 GTTTGTAAGCGACCTGGCGCAGACGCGCTGTGAATAATGCTGCCAAT 620
      ::::: |||||::: |||||::: ::::: ||
196 IleIleSerThrAspArgLeuThrLeuProThrGlnAspLeuValSerIL 212
      ::::: |||||::: |||||::: ::::: ||
621 C...GAACACATGAAATCGCGCGCGCATCTCGCGGTGAGTGCGCA 667
      ::::: |||||::: |||||::: ::::: ||
212 eAlaHISLeuHISThrIleAspILyProPheProSerGILySerProSerT 229
      ::::: |||||::: |||||::: ::::: ||
668 CGCACATTCATTCATCGACCGCGTC...GGCGCAATAAACCGGTGCG 714
      ::::: |||||::: |||||::: ::::: ||
229 hrHISLeHISLeHISLeAlaArgILeArgAsnILnArgSpValValPhe 245
      ::::: |||||::: |||||::: ::::: ||
715 ACCATCAATATCAAGATGTAATTTACCATTCGCGCTGTGTGCAACAGG 764
      ::::: |||||::: |||||::: ::::: ||
246 ThrIleSerPheGILnGILyValLeuSerILeGILyHISLeuPheLysIL 262
      ::::: |||||::: |||||::: ::::: ||
765 CGCTGTGAACACCGACGCGCTGATTCGCTAGGTGTCTCAAGTC.... 810
      ::::: |||||::: |||||::: ::::: ||
262 yPheValILeUGILyGILnGILyValAlaLeuAlaGILySerAlaLeuProp 279
      ::::: |||||::: |||||::: ::::: ||
811 ..AACAAACGCGCGCTTCGCTGACCGCTTTGGTGGCGAATATGTCGAA 858
      ::::: |||||::: |||||::: ::::: ||
279 roSerGILnArgILyThrLeuILeThrAlaLysGILyAlaSerPheSerasp 295
      ::::: |||||::: |||||::: ::::: ||
859 ATTACGCGCGGGAATGTTGTACACAGACAAAC...CGCGTATTTCCGG 905
      ::::: |||||::: |||||::: ::::: ||
296 LeuLeuProLysAspILePheSerSerAspGILyILeThrLeuILeSerIL 312
      ::::: |||||::: |||||::: ::::: ||
906 TTGCGATTTGAACGCGCGCATTCACAAAGCGCGCAGCAT...TATTGG 952
      ::::: |||||::: |||||::: ::::: ||
312 yAspProLeuThrGILyArgLeuCyLysILyGILnLysnProCySLeUG 329
      ::::: |||||::: |||||::: ::::: ||
953 GACGCTACCAATGACAGATTTCCGTTATCGAAGAGCGCGCAGCAAGAG 1002
      ::::: |||||::: |||||::: ::::: ||
329 LyMetArgSpHISThrILeThrLeuLeuProAsnProLysThrArgILn 345
      ::::: |||||::: |||||::: ::::: ||
1003 CHTGTCGGTGGTGGTGGCGCGCGACGCAACATCTCATACCGCGTAC 1052
      ::::: |||||::: |||||::: ::::: ||
346 SerPheSerPheLeuArgLeuGILyTrpAsnLysLeuThrValThrArgTh 362
      ::::: |||||::: |||||::: ::::: ||
1053 AACCCCTGGCCATTTCTGAAAAACA...CTCTTCAAGTTCACCA 1096
      ::::: |||||::: |||||::: ::::: ||
362 TrpLeuSerGILyPhePheLysArgILyArgValAlaPheMetLysPheCAsPT 379
      ::::: |||||::: |||||::: ::::: ||
1097 CAGCCGTCAGCGCGCGCAGCGCGCATGTCGCCGATTGACTATTCAGAG 1146
      ::::: |||||::: |||||::: ::::: ||
379 hrAsnMetHISGILyLysArgProIleIleAspAlaGILnILeTyILn 395
      ::::: |||||::: |||||::: ::::: ||
1147 CGCGTGAATGCGCTGTGATTCGCCACCGCTGTTGGCGCGCATTTAT 1196
      ::::: |||||::: |||||::: ::::: ||
396 ArgValSerAlaIleProValProValAlaLeuIleIleLysAlaLeuGIL 412
      ::::: |||||::: |||||::: ::::: ||
1197 CGTCGCGCATACCGACGCGCGCAGCGCATGCGTGTGGAATTTGAGCG 1246
      ::::: |||||::: |||||::: ::::: ||
412 uThrGILnAsnPheGILnLysArgLeuGILyLeuLeuGILnValAlaP 429
      ::::: |||||::: |||||::: ::::: ||
1247 AAGAAGACCTGCTTGTGCAAGCTTGTCTCCGCGCAATTCACCAATAC 1296
      ::::: |||||::: |||||::: ::::: ||

```

```

429 roGILAspPheAlaLeuProThrPheIleAspProSerLysThrGILuMet 445
1297 GCGCCGCTGTCGCAAGTGTGGAACCAATGGAAGGAA 1338
      ::::: |||||::: |||||::: ::::: ||
446 PheSerILeValLysGILSerLeuLeuArgILyAlaLysGILn 459
seq_name: SwissProt_40:RNFC_VIBCH

```

```

seq_documentation_block:
ID RNFC_VIBCH STANDARD; PRT; 774 AA.
AC O9K788;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Electron transport complex protein rnfC.
GN RNFC OR VC1015
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrrio.
OX NCBI_Taxid=665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Retchlin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
CC -1- FUNCTION: May be part of a membrane complex involved in electron
CC transport (By similarity).
CC -1- COFACTOR: Binds 2 4Fe-4S clusters (Potential).
CC -1- SUBUNIT: Composed of at least six subunits; rnfA, rnfB, rnfC,
CC rnfD, rnfE and rnfG (By similarity).
CC -1- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -1- SIMILARITY: BELONGS TO THE 4FE4S BACTERIAL-TYPE FERREDOXIN FAMILY.
CC RNFC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: AE004183; AAF94176.1; ALT_INIT.
DR TIGR: VC1015;
DR InterPro: IPR001450; 4Fe4S_ferredoxin.
DR InterPro: IPR001949; Complex1_51K.
DR Pfam: PF001512; Complex1_51K; 1.
DR Pfam: PF00037; Ier4; 2.
DR PROSITE: PS00198; 4FE4S_FERREDOXIN; 2.
KW Electron transport; Iron-sulfur; 4Fe-4S; Inner membrane;
KW Complete proteome.
KW
FT METAL 378 378 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 381 381 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 384 384 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 388 388 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 417 417 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 420 420 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 423 423 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 427 427 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 427 427 IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ SEQUENCE 774 AA; 83500 MW; 44EB13EC304E3A4E CRC64;

```

```

alignment_scores:
Quality: 158.00 Length: 472
Ratio: 0.721 Gaps: 16

```





RT electron transport to nitrogenase.";  
 RL Mol. Gen. Genet. 241:602-615(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A. / AND SEQUENCE OF 1-6.  
 RC STRAIN-ATCC 33303 / B10;  
 RX MEDLINE=98151232; PubMed=9492268;  
 RA Jouanneau Y., Jeong H.-S., Hugo N., Meyer C., Willison J.C.;  
 RT "Overexpression in Escherichia coli of the rnf genes from Rhodospirillum rubrum - characterization of two membrane-bound iron-sulfur proteins.";  
 RL Eur. J. Biochem. 251:54-64(1998).  
 RN [3]  
 RP SUBCELLULAR LOCATION.  
 RC STRAIN=SB1003 / St. Louis;  
 RX MEDLINE=9729785; PubMed=9154934;  
 RA Kumagai H., Fujiwara T., Matsubara H., Saeki K.;  
 RT "Membrane localization, topology, and mutual stabilisation of the rnfABC gene products in Rhodospirillum rubrum and implications for a new family of energy-coupling NADH oxidoreductases.";  
 RL Biochemistry 36:5509-5521(1997).  
 CC -1 FUNCTION: Required for nitrogen fixation. May be part of a membrane complex functioning as an intermediate in the electron transport to nitrogenase. Stabilizes rnfB.  
 CC -1 COFACTOR: Binds 2 4Fe-4S clusters (Potential).  
 CC -1 SUBUNIT: Composed of at least six subunits: rnfA, rnfB, rnfC, rnfD, rnfE and rnfG.  
 CC -1 SUBCELLULAR LOCATION: Inner membrane-associated (Probable).  
 CC -1 INDUCTION: Expression is reduced under iron-limiting conditions.  
 CC -1 SIMILARITY: BELONGS TO THE 4FE4S BACTERIAL-TYPE FERREDOXIN FAMILY.  
 CC RNF SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/or-send-an-email-to-license@isb-sib.ch>).  
 CC -----  
 DR EMBL: X72888; CA51399.1; -;  
 DR EMBL: Y11913; CA72670.1; -;  
 DR InterPro: IPR001450; 4Fe4S\_ferredoxin.  
 DR InterPro: IPR001949; Complex1\_51K.  
 DR Pfam: PF01512; Complex1\_51K; 1.  
 DR PROSITE: PS00198; 4Fe4S\_FERREDOXIN; 2.  
 KM Nitrogen fixation; Electron transport; Iron-sulfur; 4Fe-4S;  
 KW Inner membrane.  
 FT METAL 381 381 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 384 384 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 387 387 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 391 391 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 420 420 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 423 423 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 426 426 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 430 430 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT DOMAIN 498 503 POLY-ALA.  
 SO SEQUENCE 519 AA; 55587 MW; B5451EB2A3FA6026 CRC64;

alignment\_scores:  
 Quality: 150.00 Length: 454  
 Ratio: 0.676 Gaps: 18  
 Percent Similarity: 48.899 Percent Identity: 19.604

alignment\_block:  
 US-09-303-518d-125 x RNF\_C\_RHOCA ..  
 Align seg 1/1 to: RNF\_C\_RHOCA from: 1 to: 519

70 CCGCGCATTCGAGATCGCGTGGCGGAGATATGCGGATATCGG 119  
 |||||::: ::::: ||| ::::: ||| :::  
 42 ProAlaLeuIleArgLeuProLeu.....GlnGlnHisIleGlyAlaG1 56

120 CCGCTCGATGAAGTCAAGGAGCGAGCCGTCAGAAAAGGCCAAGTGC 169  
 : :::: |||||::: ||| ||| |||||:::  
 56 uAlaGluProIleValIysArgAspAspLeuValLeuIysGlyGlnLeuI 73  
 170 TGTGTGAAGCAAAAAGATCGGGCGGTGTCTACTGCGCGGCTTCA 219  
 :: :::: ::::: ::::: |||||::: |||  
 73 LeAlaIysAlaArgGlyProLeuSerAlaAsnIleHisAlaProHisSer 89  
 220 GCGAAATGCGCGCAT.....CACCGTGGCGGAAA 251  
 |||||::: |||||::: ||| :::  
 90 GlyArgValIleAlaValGlyHisPheValAlaProHisAlaSerGlyLe 106  
 252 GCGCGTACTGATGATCGTGTATGCGGTGAAGCAAGCAAGCA..... 297  
 ||| ::::: ||| ::|||::: |||||:::  
 106 uProVal...ProThrIleThrIleArgProAspGlyIleLysPylsTrpG 122  
 298 ..ATGAGTTGACGCTACGACCGCTGAAGCGGTGGCAAACTTAAAGCGC 345  
 ::::: ||| ||||| ::::: ||| :::  
 122 LyrHisIleuProArgLeuArgProGlu.....AsnAlaAlaPro 135  
 346 GAGAGATGCGCGCGCAACCTGATCCATCCGATTGTGTGG..... 384  
 |||||::: ::::: |||||::: |||||:::  
 136 GluGluIleAlaIleAlaGlnValAlaAlaIleGlyValGlyMetGlyL 152  
 385 .....ACTGGCGTGGCGACCGCGTCCGTTGACGAAAATTCTCTG 421  
 152 yAlaThrPheProSerAlaValIysLeuAsnLeuArgGlyAlaIysPyls 169  
 422 CCGTGGATGCGGCGCGTTCGCAATTCGTAATGCGATGACACCAAT 471  
 ::::: |||||::: |||||::: |||||:::  
 169 euThr.....ThrIleIleHisGlnGlyAlaGlyGlu 180  
 472 CCGCGTGGCGCGGACCTGATGATGCAAGACCGCGGAGATT 521  
 ||| ::::: |||||::: |||||::: |||||:::  
 181 ProTyrLeuThrCysAspArgLeuMetArgGlyAlaGluGluIle 197  
 522 CAACGCGCGCTGTGGATTGACCGCTTGAACGCAAGCAAAATTCATG 571  
 |||||::: ::::: |||||::: |||||:::  
 197 eAlaAspGlyIleGlyIleMetAlaArgAlaLeu.GlyValIysGlnVal 213  
 572 TTTGTA..... 577  
 |||||  
 214 PheValAlaIleGluSerAsnLysProGlnAlaIleGluAlaMetThr 230  
 578 .....AGCAGCTGGCGGAGAGCTGCGGTCAAAATGCTGCCAAT 620  
 ||| |||||::: ||| ::::: |||||:::  
 230 gTyrAsnArgAlaLeuGlyTyrThrPheLysIleHisValValProThr 247  
 621 CGAAACATGATTCGCGCGCGCATCTGCGGTTGAGTGGCAGCG 670  
 |||||::: |||||::: |||||::: |||||:::  
 247 LysTyrPrometGlySerGlnLysIleValLysMetIleThrGlyGln 263  
 671 ACATTCATTCATGAGCGCGCGCGCATTAACCGGTGTGAGCATC 720  
 ::::: |||||::: |||||::: |||||:::  
 264 GluThrProAlaArgAlaLeuThrAlaAspLeu.GlyValValValHis 280  
 721 AATATGACATGTAATTACCATTTGGCGTTGTTCGCAACAGCGGCTCT 770  
 ::::: |||||::: |||||::: |||||:::  
 280 snIleAlaThrAlaHisAlaValHisLeuAlaValArgTyrGlyLeuP 296  
 771 GAACCGGAGCGGTGATGCGCGTGGTGTTCAGAGTCAACAAACCG 820  
 |||||::: |||||::: |||||::: |||||:::  
 296 OleuIleAlaArgThrValThrValSerGlyHisGlyIleArgArgPro 313  
 821 GCGTTCGCTACCGTTTGGGTGCGAAGTATGCAAAATTTACTGCG... 867  
 |||||::: |||||::: |||||::: |||||:::  
 313 LAsnLeuArgValLeuIleGlyThrProValSerGluIleIleAlaHis 329  
 868 ...GGCATTGGTGCACAGACAGACCGGCTATTCGCGTTCGCTAT 914  
 ||| ::::: |||||::: |||||::: |||||:::  
 330 CysGlyGlyPheThrGluGluProAspArgLeuLeuGlyGlyLeuPro 346  
 915 GAACGCGCGATTACACAGCGCGCATTTATTTGGAGCGCTACACA 964



```

163  lueInlyValylsleuLeuilelleasnglyAlaGluCysGluProtyr 179
478  GGTGGCCGACCTACGCTCATTTATCAAGAACGCCGCGAGATTCAACG 527
180  lIethCysAspAspArgleuMetArgGluArgAlaAspGluilelle 196
528  CGGCCGTGTGTATGAGCGTTTGACCGACGCAAAATCCATGTTTGA 577
196  sgllyleArgileleuArgtyrilleuHisProGlu..LysValValI 212
578  AGCGACCTGGCGACAGCTGCCGCTC..... 603
212  leAlalleGluAspAsnLysProGluAlalleSerAlalleArgAsnAla 228
604  ...GAAATGCTGCCACATCGAACACATGATTCGGGGCGCCGACAC 650
229  leuGlnGlyAlaAsnAspIleSerileArgValIleProThrLysTyr 245
651  TCCCGGTTTGAGTGGCAGCCACATTCATTCATCGAG.....C 688
245  oserGlyAlaThrLysGlnleuiletyrleuLeuThrGlyIleGluValP 262
689  CGGTGGCGCCGAATAAACCGGTGACCATTAATTCAGATTAAT 738
262  roSerGlyGluArgSerSerileGlyValleuMetGluAsnVal... 277
739  ACCATTGGCGGTTTGTTCACACAGCGCT.....CTGAC..... 774
278  ....GlyThrMetPheAlalleLysArgAlalleleAsnAspGluPr 292
775  ...ACGAGCGCGTGAATGCCCTAGTGTTCTCAGTCAACAACCGC 820
292  oleuileGluArgValIleThrleuThrGlyAsnLysIleAlaGluLysG 309
821  GCGTTCGTGACCGCTTTGGTGGCAATGCCAAT.....ACT 864
309  lYAsnTyrTrpAlaArgleuGlyThrProIleSerGlnIleLeuSerAsp 325
865  GCGGCGAATTTGGTTCACACAGACACACCGCGATTCCGGTGGTATT 914
326  AlaGlyTyrGlnPheAspLysHisPheProIlePheAlaGlyGlyProme 342
915  GACGCGCGCGATTACACAGCGCGCAGATTATTGGAGCTACACACA 964
342  lMetGlyLeu..... 345
965  ATCAAGATTCCGTATCGAAGAGCGCGCACCAAGAGCTGTGGCTGG 1014
346  ..GluLeuProAsnleuAsnAlaProValThrLysLeuValAsnCysLeu 361
1015  GTTGGCGCGCAGCGGACCAATCTCCATCAGCGGTACACCCCTGGCGCA 1064
362  leuAlaProAspTyrleuGluTyr..... 369
1065  TTTCTGAAAAAACAATCTTCAAGTTCACACAGCGCTCAGCGCGCG 1114
370  .....AlaGluProGluAlaG 375
1115  ACCGCGCATGTGCCATTTGTACTACGAGCGCGTGAGCCTTGAT 1164
375  lGlnAlaCysileArgCysSerSerCysSerAspAlaCysProValAsn 391
1165  ATCCTGCGCACCTCTTTTGGCGATTATATCTGCGCGCTACGACAG 1214
392  leuMetProGlnGlnleuTyrTrpPheAlaArgSerGluAspHisLysLys 408
1215  CGCGCAGCATTTGGT.....TCCTTGAATTGGACGAAGACAC 1255
408  sSerGlnGluTyrAlaLeuLysAspCysIleGlu.....C 420
1256  TCGCTTTGTGAGCTGTCTGCGCGCGGCAAAATACGAATACGCGCGCTG 1305

```

```

420  ysglyIleCysAlaTyrValCysPro.....SerHisIleProLeu 433
1306  TTGGCGC.....AAAGTGTGGAACCATTTGAGAA 1334
434  lIeGlnTyrPheArgGlnGluLysAlaLysIleThrPbIleLysGluTy 450
1335  GGAA 1338
450  sGln 451

seq_name: SwissProt_40:HKRI_YEAST
seq_documentation_block:
ID      HKRI_YEAST      STANDARD:      PRT: 1802 AA.
AC      P41809;
DT      01-NOV-1995 (rel. 32, Created)
DT      01-NOV-1995 (rel. 32, Last sequence update)
DT      01-OCT-1996 (rel. 34, Last annotation update)
DE      Hansenula MRKRII killer toxin-resistant protein 1 precursor.
GN      HKRI OR YDR420W.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX      NCBI_TaxID=4932;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-YNN 295;
RX      MEDLINE=94156857; PubMed=8113191;
RA      Kasahara S., Yamada H., Mio T., Shitatori Y., Miyamoto C.,
RA      Yabe T., Nakajima T., Ichishima E., Furuchi Y.;
RT      "Cloning of the Saccharomyces cerevisiae gene whose overexpression
RT      overcomes the effects of HM-1 killer toxin, which inhibits
RT      beta-glucan synthesis."
RL      J. Bacteriol. 176:1468-1499(1994).
CC      -I- FUNCTION: COULD REGULATE BETA-GLUCAN SYNTHESIS. OVEREXPRESSION
CC      PROVIDES RESISTANCE TO HM-1 KILLER TOXIN.
CC      -I- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC      -I- PTM: COULD BE O-GLYCOSYLATED IN SERINE/THREONINE RICH DOMAIN.
CC      -I- SIMILARITY: SOME TO YEAST MSB2.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL, S69101; AAB30051.1; -.
DR      SGD: S0002826; HKRI.
KW      Glycoprotein; Transmembrane; Repeat; Signal.
FT      SIGNAL          1      21
FT      CHAIN           22      1802
FT      TRANSMEM       1486     1506
FT      DOMAIN         23      1478
FT      DOMAIN         453      788
FT      REPEAT         453      480
FT      REPEAT         481      508
FT      REPEAT         509      536
FT      REPEAT         537      564
FT      REPEAT         565      592
FT      REPEAT         593      620
FT      REPEAT         621      648
FT      REPEAT         649      676
FT      REPEAT         677      704
FT      REPEAT         705      732
FT      REPEAT         733      760
FT      REPEAT         761      788
FT      CARBOHYD       24      24
FT      CARBOHYD       1252     1252

```



|               |  |
|---------------|--|
| XX            | NCBI_TaxID=10116;  |
| RN            | [1]  |
| RP            | SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  |
| RC            | STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;   |
| RX            | MEDLINE=93328754; PubMed=8335683;  |
| RA            | Hallberg E., Wozniak R.W., Blopel G.;  |
| RT            | "An integral membrane protein of the pore membrane domain of the   |
| RL            | nuclear envelope contains a nucleoporin-like region.";   |
| J. Cell Biol. | 122:513-521(1993).   |
| CC            | -1- FUNCTION: ESSENTIAL COMPONENT OF THE NUCLEAR PORE COMPLEX. THE   |
| CC            | REPEAT-CONTAINING DOMAIN MAY BE INVOLVED IN ANCHORING COMPONENTS   |
| CC            | OF THE PORE COMPLEX TO THE PORE MEMBRANE.  |
| CC            | -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. NUCLEAR PORE   |
| CC            | MEMBRANE.  |
| CC            | -1- DOMAIN: CONTAINS X-F-X-F-G REPEATS.  |
| CC            | -1- SIMILARITY: THE REPEAT REGION COMPOSED OF PENTAPEPTIDE REPEATS   |
| CC            | SEPARATED BY SER/THR-RICH DOMAINS IS SIMILAR TO THAT OF YEAST  |
| CC            | NOP1, NSP1 AND MAMMALIAN P62 AND NOP153.   |
| CC            | -----  |
| CC            | This SWISS-PROT entry is copyright. It is produced through a collaboration   |
| CC            | between the Swiss Institute of Bioinformatics and the EMBL outstation  |
| CC            | the European Bioinformatics Institute. There are no restrictions on its  |
| CC            | use by non-profit institutions as long as its content is in no way   |
| CC            | modified and this statement is not removed. Usage by and for commercial  |
| CC            | entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> |
| CC            | or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).  |
| CC            | -----  |
| DR            | EMBL; Z21513; CAAV9725.1; -  |
| KW            | EMBL; Z21514; CAAV9726.1; -  |
| RN            | Nuclear protein; Transport; Transmembrane; Repeat.   |
| FT            | DOMAIN 1 56  |
| FT            | DOMAIN 57 77   |
| FT            | DOMAIN 76 1199   |
| FT            | DOMAIN 4 8   |
| FT            | DOMAIN 53 58   |
| FT            | DOMAIN 438 441   |
| FT            | DOMAIN 497 500   |
| FT            | DOMAIN 684 687   |
| FT            | DOMAIN 1023 1026   |
| FT            | DOMAIN 1033 1038   |
| FT            | DOMAIN 1058 1063   |
| FT            | DOMAIN 1199 AA; 120784 MW; 60D4451B91D5B907 CRC64;   |
| SS            | SEQUENCE   |
| SS            | 1199 AA; 120784 MW; 60D4451B91D5B907 CRC64;  |

```

alignment_scores:
  Quality: 129.50
  Ratio: 0.568
  Percent Similarity: 47.799
  Length: 477
  Gaps: 1
  Percent Identity: 21.38

```

```
alignment_block:
US-09-303-518D-125/rev x NI21_RAT .
```

Align seg 1/1 to: NI21\_RAT from: 1 to: 1199

[illegible][illegible]

295 CCGTCGTCCTTACAGCGCATACAGCTGAGTACAGCGCTTTTGG 246  
 :|||||  
 868 LaserSerAlaGlnThrAla..... 874  
 245 CCAAGCGTAAATCCGGGATTTTGCCTGAAGCGCGCAGTAACACAC 196  
 :|||||  
 875 .....AlaSerSnsrSer 879  
 195 GCCCGGATTC.....TTTGTGCTTCAACAGCAGCTTGGC 161  
 :|||||  
 879 TGIyGlyPheSerGlyPheGlyGlyThrLeuThrThrSerThrSerAlaP 896  
 160 CTTTGTTCAGCGGATGCTTCCTTCCTTCATGCGAGGCGCATACCG 111  
 :|||||  
 896 roAlaThrThrSerGlnProThrLeuThrPheSerSerThrValThrPro 912  
 110 GCATATTCCTTCCCAAGCAAGCGACTTCGTAATAGCGCGCGCTGTA 61  
 :|||||  
 913 ThrPheSnlleProPheSerAlaSerAlaLysProAlaLeuProThrTy 929  
 60 AACGCGTTCCTCGGCTCGCGCGGATGGC 30  
 :|||||  
 929 rProGlyAlaSnsrGlnProThrPheGly 939

seq\_name: SwissProt\_40:RNFC\_PSEAE

seq\_documentation\_block: PRT: 774 AA.

ID RNFC\_PSEAE STANDARD: PRT: 774 AA.  
 AC 09HYB8;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DE 01-MAR-2002 (Rel. 41, Last annotation update)  
 GN Electron transport complex protein rnfC.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 ON NCBI\_TaxID=287;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PAOI;  
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 Brody L.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yan Y.,  
 Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT Complete genome sequence of Pseudomonas aeruginosa PAOI, an  
 opportunistic pathogen.  
 RL Nature 406:959-964(2000).  
 CC - FUNCTION: May be part of a membrane complex involved in electron  
 transport (By similarity).  
 CC - COFACTOR: Binds 2 4Fe-4S clusters (Potential).  
 CC - SUBUNIT: Composed of at least six subunits: rnfA, rnfB, rnfC,  
 rnfD, rnfE and rnfG (By similarity).  
 CC - SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).  
 CC - SIMILARITY: BELONGS TO THE 4FE4S BACTERIAL-TYPE FERREDOXIN FAMILY.  
 CC RNFC SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: AE004770; AAC06879.1;  
 DR InterPro: IPR001450; 4Fe4S\_Ferredoxin.  
 DR InterPro: IPR001949; Complex1\_51K.  
 DR Pfam: PF01512; Complex1\_51K; 1.  
 DR Pfam: PF00037; fer4; 2.

DR PROSITE: PS00198; 4FE4S\_FERREDOXIN; 2.  
 KW Electron transport; Iron-sulfur; 4Fe-4S; Inner membrane;  
 KM Complete proteome.  
 FT METAL 369 369 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 372 372 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 375 375 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 379 379 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 408 408 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 411 411 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 414 414 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 418 418 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 SQ SEQUENCE 774 AA; 82359 MW; C82B06F4D2E09B6 CRC64;

alignment\_scores: Length: 488  
 Quality: 129.00 Gaps: 23  
 Ratio: 0.578  
 Percent Similarity: 45.697 Percent Identity: 21.516

alignment\_block: US-09-303-518D-125 x RNFC\_PSEAE ..

Align seg 1/1 to: RNFC\_PSEAE from: 1 to: 774

28 CCGCCATCGCGCGCGGCGAGCGAGCGAGCGCGGCGGCGCAT 77  
 :|||||  
 23 LeuProIleGlnGlnAlaProLeuAlaGlnGlyThrIleValPro.... 37  
 78 TACCGAAGTCGCGGTCCTTGGCGAAGATAT...GCCGTAATGCCCGCCT 124  
 :|||||  
 38 .....LeuGlyGlnHisIleGlyAlaProAlaValProC 49  
 125 CGATGAATCAAGGAGGCGATGCCGTCGTAAGGCGCAAGCTGCTTT 174  
 :|||||  
 49 yValGlnVal.....GlyGlnAlaValLeuLysGlyThrIleAla 63  
 175 GAAGCAAAAGAAATCCGCGCGGCTTACTGCGCGGCTTCAGGCA 224  
 :|||||  
 64 LeuProAspGlyThrValSerAlaIleLeuHisAlaProThrSerClyTh 80  
 225 AATCGCGCGGAT.....CACCGTGGCGAAAGCGCGTAC 259  
 :|||||  
 80 yAlaValAlaIleGlyAlaHisProGlyProHisAlaSerGlyLeuProA 97  
 260 TTCAGTCAGTCGATGTCGCGTGAAGCAAGCAGGAG..... 297  
 :|||||  
 97 LeuAlaIleValIleAlaSerAspGlyLeuGlnGlyThrIleLeu 113  
 298 .....ATCGAGTTGAAGCGTACGACCTGAAGCGCTGCAAACTT 338  
 :|||||  
 114 HisProCysProAspPheArgAlaGlnSerProLeuAlaLeu..... 128  
 339 AACGCGCGAAGATCGCGCC.....AACGTATCCAAATCG 376  
 :|||||  
 129 .....GluArgIleArgAlaAlaGlyIleGlyGlyLeuGlyAlaG 143  
 377 GTTGTGGAGCTGGC.....CTGCGCACCGTCGCTGCAAAATTCCT 420  
 :|||||  
 143 LysPheProThrAlaAlaLysLeuAlaAlaArgProAlaLysIle... 158  
 421 GCCGTGATGCCGAGCGGTCGCGCATCTTGTCATATGCGATGACCA 470  
 :|||||  
 159 .....HisThrLeuValValAsnGlyAlaGlnCysG 169  
 471 TCGCGTCGCGCGCGACCTACGTCATTTTCAAGAGCGCGGAGAT 520  
 :|||||  
 169 uProTyrIleSerAlaAspAspLeuLeuLysGlnArgAlaThrGlnV 186  
 521 TCAACGCGCGCTGTGTTGATGAGCGT..... 549  
 :|||||  
 186 alleuGlyGlyIleAspIleLeuValGlnIleLeuGlyProGlnGlnVal 202  
 550 .....TTGACCGAAGCAAAATTCATGTTTGAAGGACAGCTGGCG 590



```

203 LeuValIgLylegluAspAspLysProGluAlaIleAlaIleAlaLeuGlyI 219
591 A.....GACGTGCGCTCTAATAATGTCGCAACATGCAACACATG 631
219 AlaLeuGlyLysLysProTyrTrpIleValAlaIleLeuProThrArgTyrP 236
632 AATTGCGCGCGCCGCAT..... 648
236 roserGlyGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLys 252
649 CTTGCC...GGTTGAGTGGCAGCAGCATTTTCATGACCGCGTGG 695
253 ProAlaAspLysLeuProAlaAspIleGlyLysLysLysLysLysLysLys 269
696 GCGGAATAAACCGTGGACCATCATATATGATGATGATGATGATGATGATGATG 745
269 YThrLeuAlaAlaVal.....HisAspAlaValIleGlu 281
746 GCGGTTGTTGGACACGCGCTGACACGCGCGTGGTGGTGGTGGTGGTGGTGGT 795
281 LysArgProLeuIleSer.....ArgIleThrLeu 291
796 GGGGTTTTCAGTCAACAACCGCGCTTGGTGGTGGTGGTGGTGGTGGTGGTGGT 845
292 AlaGlyGlyAlaLeuGlyLysLysLysLysLysLysLysLysLysLysLys 308
846 GAAAGTATGCAATTTACTGCG.....GGCAATGG 877
308 ProValHisGlyLeuLeuAlaPheAlaGlyLeuAlaGlyLeuValGlyLeu 324
878 TTGACACAGACAAACCGGTGATTTCCGTTGGTGGTGGTGGTGGTGGTGGTGGT 927
325 .....GluArgValLeuMetGlyGly..... 331
928 ACACAAGCGCGCAGATTATTGGAGCGGTACACACATGATTTCCGT 977
331 ..... 331
978 TATCGAAGAAAGCGCGCAGCAAGACCTGTTGCGTGGTGGTGGTGGTGGTGGT 1027
332 .....PrometMetGlyPheAlaLeuProAsp. 340
1028 CGGACAAATATCTCCATCAACGCGGTACAAACCTTGGCCATTTCTGAAAAAC 1077
341 .....LeuSerVal 343
1078 AAACCTTTCAAGTTCAACACAGCGGTCAACGCGCGCGCGCGCATGTTATTC 1127
344 ProLeuIleLysThrCysAsnCysLeuLeuAlaGlyLysPheAlaThrGlyLe 360
1128 GCGG.....ATTGTAATTACGAGC 1147
360 ProGluProValProAlaMetProCysIleArgCysGlyAspCysAlaG 377
1148 GCGTATGCGCTTGGATATCTGCCCAACCGCTTGGCCGATTTATTC 1197
377 InValCysProValSerLeuLeuProGln...GlnLeuHisPheAla 392
1198 GTGCGCGATTCAGACAGCGCGCGCATTTGGTGGTGGTGGTGGTGGTGGTGGT 1247
393 LeuGlyLys...GlnHisGlyLysLeuLeuAlaHisAsnLeuPheAsp 408
1248 AGAAGACCGCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1297
408 sIleGlyCysGlyAlaCysAlaTyrValCysProSerSerIle..... 422
1298 GCGCGCTGTGGCG 1311
423 ..ProLeuValGln 426
seq_name: SwissProt_40:RNFC_BUCAI

```

```

seq_documentation_block:
ID RNFC_BUCAI STANDARD: PRT: 473 AA.
AC P57215;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Electron transport complex protein rnfc.
GN RNFC OR B0115.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_Taxid=118099;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=TORO 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Matanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
Buchnera sp. Aps.";
RL Nature 407:81-86(2000).
CC -! FUNCTION: May be part of a membrane complex involved in electron
transport (by similarity).
CC -! COFACTOR: Binds 2 4Fe-4S clusters (Potential).
CC -! SUBUNIT: Composed of at least six subunits; rnfa, rnfb, rnfc,
rnfd, rnfe and rnfg (by similarity).
CC -! SUBCELLULAR LOCATION: Membrane-associated (by similarity).
CC -! SIMILARITY: BELONGS TO THE 4FE4S BACTERIAL-TYPE FERREDOXIN FAMILY.
CC RNFC SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use. By non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb.ch/announce/
or send an email to license@isb.slb.ch)
-----
DR EMBL: AP001118; BAB12833.1; -
DR InterPro: IPR001450; 4Fe4S_Ferredoxin.
DR InterPro: IPR001949; Complex1_51k.
DR Pfam: PF001512; Complex1_51k; 1.
DR Pfam: PF00037; fer4; 2.
DR PROSITE: PS00198; 4FE4S_FERREDOXIN; 2.
KW Electron transport; iron-sulfur; 4Fe-4S; Complete proteome.
FT METAL 337
FT METAL 337
FT METAL 340
FT METAL 343
FT METAL 343
FT METAL 347
FT METAL 347
FT METAL 376
FT METAL 379
FT METAL 382
FT METAL 382
FT METAL 386
FT METAL 386
SQ SEQUENCE 473 AA; 54563 MW; 04B64102F1315AEE CRC64;
-----
alignment_scores:
Quality: 126.00 Length: 449
Ratio: 0.575 Gaps: 22
Percent Similarity: 48.775 Percent Identity: 20.045
alignment_block:
US-09-303-518D-125 x RNFC_BUCAI ..
Align seg 1/1 to: RNFC_BUCAI from: 1 to: 473
-----
100 GAAAGAAATATCGCGTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 149
.....
13 AspAspTyrHisAsnValLys.....LeuArgValLysIleAsnGlyLys 27
-----
150 CGTCAAAAGGCGCA...GTGCTGTTGAAGCAAAAGATCCGCGCGCG 196
|||||
27 sValLeuArgGlyGlnProLeuIlePheSerAspAspPheAsn.....Y 42

```

```

197 TGGTGTACTGCGCGGCTTCAAGCAAAATCGCCCGGATTCACCGTGGC 246
198 |||||
42 alProValHisAlaProThrSerGlyLeuIleGluAsnIleCysPheAsn 58
247 GAAAGCGCGCTACTGATCGATCGATGCGCGGATTCGCGTGAAGGC..... 288
248 |||||
59 SerAspSerIleGlyAsnIleGlyIleValIleSerProAspTyrIle 75
289 .....AACGACGAAATCGAGTTTGAACGCTACGAC 319
75 uAspGlnTrpIleArgLeuAsnProIleGlyAspTyrIleGlyValAla 92
320 CTGAGCGCGTGGCAAACTTAAGCGCGCAAGAAAGTGGCCGCAACCGTATC 369
321 |||||
92 roGluIleLeuIleGlyIle.....IleHis 100
370 CAATCGCGTTTGTGACTGCGTGGCGACCGCGTCCGTTTC.....AG 410
371 |||||
101 GlnSerIleGlyAlaValGlyLeuGlyGlyGlnPheProSerIle 116
411 CAAATTCCTGCGCGTGCATGCGCGCGTCCGATCTTGTCAATGCCA 460
412 |||||
116 slyIleIleIlePheSerIleAsnArgAlaHisThrLeuIleValAsnAla 133
461 TGGACACCAATCCGCTGCTGCCGACCGCATCGGTCATTCAAAGAGCC 510
462 |||||
133 alGluSerGluProTyrIleThrSerAspAsnCysLeuIleTyrAsnHis 149
511 GCCGAGATTCGCAAGCGCGCTGTTGTATGAGCCGTTTGACCGCAAGC 560
512 |||||
150 IleSerGluIleLeuIleGlyCysIleIleLeuGlyTrpIleThrIle 166
561 CAAATTCATGTTTGTAGGACAGCTGGCGCGACAGCTCCGCT..... 603
562 |||||
166 elyThrValLeuIleAlaIleGlnIleGluAspAsnIleGlnSerIle 183
604 .....CAAAATGCTGCCACATCGCAACACATGAAATTC 636
183 yslIleGlnHisLeuIleGlyAsnIleSerLeuPheIleCysIleIle 199
637 GCGCGCGCGCATCTCCGCT.....TTGAG 662
200 LysIleIleGlyTyrProIleGlySerSerIleValLeuValIleGly 216
663 TGGCAGCGACAT.....CATTCATGCGCGCGTGGCGCGCA 700
216 rGlyIleValIleProHisGlyIleHisSerIleAsp.....IleGly 230
701 ATAAACCGTGTGACCATCATATATCAAGATGTAATTACATGGCCGT 750
231 .....TyrLeuIlePheAsnValAlaIleThrIlePheSerIleGly 244
751 TTGTTTCAACAGCGCGCTGTGAACCGCGCGGATTCGCGCAAGTGG 800
245 AlaIleIleAsnGlyLysProLeuThrGluIleValIleThrLeu..... 259
801 TTCTCAAGCAACAAACCGCGCTCTGCTACCGTTTGGGTGGCAAG 850
260 .....MetSerAspLysAsnLeuLeuSerGlyAsnPheTrpValAla 274
851 TATGCGCA.....ATTACTGGCGCGCAATTTGGTGGACCA 885
274 leGlyThrProIleGlyTyrPheLeuThrSerAsnLysLeuGlyGln 290
886 GACACCGCGGTATTCGGTGGTATTCGCGCGGATTCACCAAG 935
291 .....PheIleIleSerValIleGlyGly..... 299
936 CGCGACGATTTATTTGGAGCGCTACCAACATGATTCGCGTATCGAAG 985
300 .....ProPheMetGlyLysIleAsnAsnLeuAsn..... 310
986 AAGCGCGCGCAAGAGCTGTGCGGTGGCGCGCGCGACCGGACAA 1035

```

```

310 ..... 310
1036 TACTCCATACCGCGCTACACCGCTGGCATTTCTGAAACAAACACTT 1085
311 HisSerIleLeuLysIleThrAsnSerIlePheIleThrIleHisLysG 327
1086 CAAGTTCAACACAGCGCGTCAACGCGCGGACCGCGCATGTCGCGATG 1135
327 uLys.....AsnGluSerIleSer.....GluLysThrCysIleArgCys 341
1136 GTACTTACAGCGCGGTATGCCCTTGATATTCCTGCCACCTGCTT... 1182
341 LysCysSerTyrValCysProValAsnLeuLeuProGlnIleGlnTyr 357
1183 .....TTGCGGATTTATGCTGGCGGATTCACGCGCGGACGAT 1226
358 TrpTyrIleLysAsnLysAsnHisValGlnThrLysIleTyrVal 374
374 uAspCysIleGlu.....CysLysAlaCysGlnLysValC 386
1227 GCGTTCGTTGAAATGAGCAAGAACCGCTGCTTGTGACGCTTCGCT 1276
1277 GCCCGGCAAAATGCAATACGCGCGCTGTTGCGCAAGTCTGGA 1323
386 yAspPro.....SerTyrIleProLeuValLysTyrPheIleGln 398
seq_name: SwissProt_40:SGS3_DROER
seq_documentation_block:
ID SGS3_DROER STANDARD; PRT; 328 AA.
AC P13730;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Salivary glue protein Sgs-3 precursor.
GN SGS3.
OS Drosophila erecta (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7220;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88332966; PubMed=3138416;
RA Martin C.H., Mayeda C.A., Meyerowitz E.M.;
RT "Evolution and expression of the Sgs-3 glue gene of Drosophila.";
RL J. Mol. Biol. 201:273-287(1988).
CC -!- DEVELOPMENTAL STAGE: PRODUCED BY THIRD-INSTAR LARVAE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL, M14652; -; NOT ANNOTATED_CDS.
CC PIR: S01359; S01359.
CC FlyBase: FBgn0012268; Dere\SGS3.
CC Repeat: signal.
CC FT SIGNAL 1 23
CC CHAIN 24 328 SALIVARY GLUE PROTEIN SGS-3.
CC SEQUENCE 328 AA; 36355 MW; 62F27F188C0F8272 CkC64;
alignment_scores:
Quality: 125.00 Length: 267
Ratio: 1.000 Gaps: 12
Percent Similarity: 46.816 Percent Identity: 25.468
alignment_block:

```

US-09-303-518D-125 x SGS3\_DROER ..

Align seg 1/1 to: SGS3\_DROER from: 1 to: 328

```

29 TGGCCATCGCGGAGACCGGAGAACCGCTTACGACGGCCGG..... 73
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
28 CysProLysArgThr...ThrProLysProCysThrThrAlaArgProTh 43
74 .....CCATTACCGACGCGGTGCTGGCGAAGATATGCGCGTAT 116
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
43 rGysAlaProValThrThrThrCysArgProProThrThrThrArgC 60
117 GCGCCCTCGATGAAGTCAAGAGGCGCATCGTCAAAAAAGGCCAAG 166
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
60 ysrProProThrThrThrThrArgCysProProThrThrArgProAla 76
167 TGCTGTTTGAAGCAAAAGATCCGGGCGTGTACTGCGCCG... 214
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
77 CysThrAlaThrThrLysArgProThrAlaArgProThrThrArgArg 93
215 ...CTTCAGCGCAAAATCGCCGATTCACCGTGGCGAAGCGGCTACT 260
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
93 rThrValArgAlaThrThrLysArgAlaThrThrArgArgThrThrLys 110
261 TCAGTCAGTCG...TGATTGCCGTTGAAGCGACAGCAAGAA...TCGAGT 304
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
110 rGAlaThrThrArgArgThrThrValArgAlaThrThrLysArgAla 126
305 TGAAGCGTACGACACCTGAAGCGCTGGCAACT...TAAGCGCGCAAGA 351
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
127 ThrArgArgThrThrThrLysArgAlaProThrArgArgThrThrLys 143
352 GTGCGCCGCAACCTGATCCATCCGCTT...TGTGAGCTGCGCTCG 395
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
143 sArgAlaThrThrArgArgAsnProThrArgArgThrThrThrArgArg 160
396 CACCGCGTCCGTCACGAAATTCCTGCGCTGCATCCGACCGCTGCGCA 445
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
160 lAprThrLysArgAlaThrThrLysArgAlaThrThrArgArgAsnPro 176
446 TCT.....TCGTCAATGCGATGACACCAATCCGC..... 475
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
177 ThrLysArgLysThrThrArgArgThrThrValArgAlaThrLysThr 193
476 .....TGGTGGCGCACCTTACGCTATTAACAAGA 506
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
193 rLysArgAlaThrThrLysArgAlaProThrLysArgAlaThrThrLys 210
507 AGCGCGCG.....AGGATTCACAAAGCGGCGCTGTGTATGAGCCGT 550
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
210 rGAlaProThrLysArgAlaThrThrLysArgAla..... 221
551 TGACGCAAGCAAAATCCATGTTGTAGCAGCGTGGCGCAGAGCTGCG 600
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
222 ...ProThrLysArgAlaThrThrLysArgAlaProThrLysArgAla 237
601 TCTGAATGCTGCCACATGAAACACATGAATTCGGCGCCGCAATCC 650
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
237 rThrLysArgAlaProThrLysArgAlaThrThrLysArgAlaProThr 254
651 TGCGCGTTGATGGCAGCG..... 670
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
254 ysaArgAlaThrThrLysArgAlaProThrLysArgAlaThrThrLys 270
671 ...ACATTCAATTCATCGAGCGGTCGGCGCAATAAAGCGTGGAGC 717
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
271 AlaThrAlaArgProThrSerLysProCysLysProCysGlyPro 287
718 A 718
287 o 287

```

seq\_name: SwissProt\_40:YM96\_YEAST

seq\_documentation\_block:

```

ID   YM96_YEAST          STANDARD;          PRT: 1140 AA.
AC   004893;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Hypothetical 113.1 kDa protein in PRES-FET4 intergenic region.
GN   YMR317W OR YMR924.09.
OS   Saccharomyces cerevisiae (Baker's Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SZ880 / AB972;
RA   Churcher C.M., Louis E.J., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.
CC   -1- DOMAIN: CONTAINS MANY SER/THR-RICH DOMAIN AND REPEATS.
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@sib-sib.ch).
CC
DR   EMBL; Z51414; CAA90835.1; -.
DR   SGD; S0004936; YMR317W.
KW   Hypothetical protein; Repeat.
SQ   SEQUENCE 1140 AA; 113070 MW; 0153EBCA24FE5427 CRC64;

```

## alignment\_scores:

```

Quality: 124.00      Length: 466
Ratio: 0.588         Gaps: 21
Percent Similarity: 45.279      Percent Identity: 23.176

```

## alignment\_block:

US-09-303-518D-125/rev x YM96\_YEAST ..

Align seg 1/1 to: YM96\_YEAST from: 1 to: 1140

```

1331 TCAATGTTCCAGACTTGGCCACAGCGGCCGCTATTCGATTGGCC 1282
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
386 SerSerThrSerSerValSerSerGluAlaProSerSer..... 399
1281 CGGGCAGACGACGCTGCACAAAGCGAGTCTTCTGTCACATTCGACG 1232
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
400 .....ThrSerSerValSerSerGluAla 408
1231 AACCCATGCTCGCGCTGCGTATCGCCAGCATTAATCGCCGCAAA 1182
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
408 lAprSerSerThrSerSerValSerSerGluAlaProSerSerThrLys 424
1181 AGCAGGTGGGAGATATCCAGGCGATCAGCGCTGCTAGTACCAAT 1132
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
425 SerSerVal.....MetSerSerGluValSerSerAla.....Th 436
1131 CGGCACCATGCGCGGTGCGCGCGCTTGACGCGCTGTGTAACCTGAGA 1082
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
436 rSerSerLeuValSerSerGluAlaProSerAlaIleSerSerLeuAla 453
1081 GT.....TTGTTTTCAGAAATGCGCCAGAGGTTGTACGCTGATG... 1041
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
453 rSerArgLeuPheSerSerLysAsnThrSerValThrSerThrLeuVal 469
1040 .....GAGTATTTCGCGGCTCGCGGCGCAACCCAGCAAGAGCTCTT 997
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
470 AlaThrGluAlaSerSerValThrSerSerLeuAlaArgProSerSer 486
996 GCTGCGGCTTCTTCGATAACGAAATCGATTGATGTCGCTCCCAAT 947

```

```

486 rleuAlaSerAsnSerIleIleGlu..... 494
494 ..... 494
896 ACGGGTGTGTGTGTCACCAATTCGCGGCAATAATTCGATCTTT 847
896 AATCGTGGCGCCTTGTGTATCGCGCGTTCAATACCGAACGGAAATC 897
494 ..... 494
896 ACGGGTGTGTGTGTCACCAATTCGCGGCAATAATTCGATCTTT 847
495 SerSerLeuSerThrGlyTyrAsnSerThrValSerThrThrSerAl 511
846 CGCACCCAAACGGTACGCAAGCGCGGTTTGTACTTGAGAACCAC 797
511 aAlaSerSerThrLeuGlySerLys..... 519
796 CTAGGGCAATCACGCGCTGCTTCAGACGCGCTTTCGAAACAAACG 747
520 ..... ValSerSerSerAsnSerAlaGlyMetAlaThrSerLysThr 532
746 CCAATGTAATACATCTTGAATTAATGATGTCACACGGTTTATTCGC 697
533 SerSer.....ThrSerAspLeuSerLysSerSerValIlePheG1 547
696 CGCGGCGGCTCGATGAAATGATGCGTGCACCTCAACCGCGAGGAT 647
547 YAsnSerSerThrVal.....T 553
646 GCGGCGCGCGCAATTCATGTTTCGATGTTGCAGCA..... 609
553 hThrSerProSerAlaSerIleSerLeuThrAlaSerProLeuProSer 569
608 .....TTTTCAGACGCGACGCTGCGCGCGCTTTCACAAACATGATTTT 562
570 ValTTPSerAspIleThrSerSerGluAlaSerSerIleSerSerAsnLe 586
561 CGGTTGCGTCAACGCGCTCAATACCAACGCGCGCTTTCGAAATCCTCGG 512
586 uAlaSerSerSerAlaProSerAspAsnAsnSerThrIleAlaSer..... 601
511 CGGCTTCTTTGATTAATGACC.....GTGGGCTCGGCA 480
602 ..AlaSerIleuIleValThrLysThrLysAsnSerValValSerSerIle 617
479 GCCACGCGATGTTGTCCATTCGATTCAGCAAGATGCGACGCGCTCGGC 430
618 ValSerSerIleThrSerSerGluThrLysAsnGluSerAsnLeuAlaTh 634
429 ATCGACGCGCAGGAAATTTTGTGACGACGAGCGGTGCGCAGCGCAGTCA 380
634 rSerSerThrSerLeuLeuSerAsn..... 642
379 AACCGGATTTGATCAGTTGGGGCGGCGCTTTCGCGCTTAAGTTTGC 330
643 .....LysAlaThrAlaArgSerLeuSerThrSer 652
329 ACGCGTTGAGTGCCTAGCGTTCAAC.....TCGATTTGCTC 292
653 AsnAlaThrSerAla.....SerAsnValProThrGlyThrPheSerSe 667
291 GTTGCCTTCA..ACGCAATCAGACATGATGAAAGTACGGCGTTTTCGC 245
667 rMetSerSerIleThrSerVal..... 674
244 CACGGTGAATCGCGCGATTTTGCCTGAAGCGCGCAGTAACACACAG 195
675 .....IleThr 676
194 CCGGATTTCTTTTGTCTTCA.....AACACACTTGGCC 160
677 ProGlyPheSerThrSerSerAlaSerLeuAlaIleAsnSerThr..... 691
159 TTTTTCGACGCGATCGCTTCTTACCTTTATTCGAGGGGCGCATACCGG 110

```

```

692 .....ValValSerSerSerLeuAlaGlyTyrSerPheSerThrProG 706
109 CATATTCCTTCGCCAAGCAACGCGACTTGCTATGATCGCGCGCTGCTGA 60
706 lu...SerSerProThrThrSerThrLeuValThrSer..... 717
59 ACGGCTGTGCTGCGTGGCGGATGCGGCGAGCTTTCAGACCTTTTTCG 12
718 GluAlaProSerThrValSerSerMetThrThrSerAlaProPheIle 733
seq_name: SwissProt_40:RNFC_ECO57
seq_documentation_block:
ID RNFC_ECO57 STANDARD; PRT; 740 AA.
AC P58324;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE Electron transport complex protein rnfc.
GN RNFC OR Z2636 OR ECG2338.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Rose D.J., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
RA Perna N.T., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimlantia E.T., Potamouis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Toke T.,
RA Iida T., Takami H., Honda T., Sasaki G., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -!- FUNCTION: May be part of a membrane complex involved in electron
CC transport (By similarity).
CC -!- COFACTOR: Binds 2 4Fe-4S clusters (Potential).
CC -!- SUBUNIT: Composed of at least six subunits; rnfa, rnfb, rnfc,
CC rnfd, rnfe and rnfg (By similarity).
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -!- SIMILARITY: BELONGS TO THE 4FE4S BACTERIAL-TYPE FERREDOXIN FAMILY.
CC RNFC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AE005386; AAG56618.1; -
CC DR EMBL: AP002558; BAB35761.1; -
CC PROSITE: PS00198; 4FE4S_FERREDOXIN; 2.
CC Electron transport; Iron-sulfur; 4Fe-4S; Inner membrane.
KW METAL
FT 377 377 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 380 380 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 383 383 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 387 387 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 416 416 IRON-SULFUR (4FE-4S) (POTENTIAL).

```





|    |  |
|----|--|
| CC | -----  |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration   |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation  |
| CC | the European Bioinformatics Institute. There are no restrictions on its  |
| CC | use by non-profit institutions as long as its content is in no way   |
| CC | modified and this statement is not removed. Usage by and for commercial  |
| CC | entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> |
| CC | or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).  |
| CC | -----  |
| DR | EMBL; AJ248283; CAB49021.1; -  |
| DR | HSSP; P22983; IDIK.  |
| DR | InterPro: IPR000121; PEP_utilizers.  |
| DR | InterPro: IPR002192; PPK_N_term.   |
| DR | Pfam; PF00391; PEP_utilizers; 1.   |
| DR | Pfam; PF02896; PEP_utilizers; 1.   |
| DR | Pfam; PF01326; PPK_N_term; 1.  |
| DR | ProDom; PD000940; PEP_utilizers; 1.  |
| DR | PROSITE; PS00370; PEP_ENZIMES; PHOS_SITE; 1.   |
| DR | PROSITE; PS00742; PEP_ENZYMES_2; 1.  |
| KW | Transferase; Kinase; ATP-binding; Phosphorylation; Complete proteome.  |
| FT | MOD_RES 441 441  |
| FT | DOMAIN 809 816   |
| FT | POSSPHORYLATION (BY SIMILARITY).   |
| FT | POXY-GLU   |
| SQ | SEQUENCE 819 AA; 90652 MW; E09486AB527C6EA CRC64;  |

|                     |              |
|---------------------|--------------|
| alignment_scores:   |              |
| Quality:            | 113.50       |
| Ratio:              | 0.473        |
| Percent Similarity: | 52.747       |
|                     |              |
|                     | Length: 4555 |
|                     | Gaps: 21     |
| Percent Identity:   | 21.538       |

alignment\_block:

US-09-303-518D-125 x PPSA\_PYRAE

Align seg 1/1 to: PPSA\_PYRAB from: 1 to: 819

[illegible][illegible]



seq\_name: SwissProt\_40:VL2\_HPV70

seq\_documentation\_block:

ID VL2\_HPV70 STANDARD; PRT; 466 AA.

AC P50801;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE Minor capsid protein L2.

GN L2.

OS Human papillomavirus type 70.

OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;

OC Papillomavirus.

OX NCBI\_Taxid=39457;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96249586; PubMed=8815087;

RA Forstlund O., Hansson B.G.; 70 genome cloned from overlapping PCR

RT products: complete nucleotide sequence and genomic organization.";

RL J. Clin. Microbiol. 34:802-809(1996).

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

alignment\_scores:

Quality: 112.50 Length: 402

Ratio: 0.349 Gaps: 21

Percent Similarity: 50.995 Percent Identity: 24.876

alignment\_block:

US-09-303-518D-125 x VL2\_HPV70 ..

Align seg 1/1 to: VL2\_HPV70 from: 1 to: 466

28 CTGCCCATCGCGGCGAGACCGGAGACCGCTTTAGACGCGCGGCCAT 77

72 ILeProLeuGlyGlyArgProSerThrValAlaSerValThrProAlaAr 88

78 TACCGAAGTCGCGTCTGGCGAAGAAATGCCGATMGCGGCCCTGGA 127

88 gProProValValIle.....GluProValJlyProThrGluProSerI 103

128 TG.....AAAGTCAAGAAAGGAGATGCCGTCAAAAAGGCCAAGTGTG 171

103 IeValGlnIleValGlnGluSerSerValValSerSercly.....I 116

172 TTTGAAGACAAAAGAAATCCGGCGGTGTACTGCGCGCGCTTCAG 221

117 .....ThProIleProThrProThrGlyThrSerclyp 128

222 CAAAATCGCGCGGATTCACCGTGGCGAAGGCGGACTTCAGTCAATG 271

128 eguIleThrSerSerAlaThrThrThrProAlaValIleuAsp.....I 143

272 TGAATCGCTTGAAGCAAGCAAGAAATGATTTGAACGCTACGACCT 321

143 IeThrProAlaSerGlySerValGlnIleSerThrThrSerThrThrAsn 159

322 GAAGCGCTGGCAAACTTAAAGCGGCGAAGAAAGTCCGCCACCTGATCA 371

116 ProAlaPheAlaAspProSerLeuIleGluValProGlnThrGlyGluVal 176

372 ATCCGGTTTGTGACGTGCGTGGCGACCCCGCGCTTGAAGCAAAATCTG 421

176 IserGly.....AsnIlePheValThrThrProThrSerclyThrIsc 191

422 CGCGTACGCGGCGCGGCGTGGCGCATCTTCGTC.....AATGCGATGACAC 468

191 IYTYGlnGlnIleProMetGlnValPheAlaSerHisGlyThrGlyThr 207

469 AATCGCTGCGTGGCGCGACCTTACGCTGATTTATCAAGAAAGCGCGCGA 518

208 GluProIleSerSerThrProValProGlyValSerAlaGlyVal.. 223

519 TTTCAAGCGCGCGCTGTGATTTAGCGCGTGGCGCAAGCAAAATCC 568

224 .....ProArgLeuThrSerArgAlaThr 232

569 ATGTTTGAAGCGACGTGCGCGACGTGCGCTTGAAGCAAAATCTGCAAC 618

232 IeGlnValArgValAsnAsnPheAspPheValThrArgProSerSerPhe 248

619 ATCGAAGCAATGATTCGGCGCGCG.....CATCGCGCGGTTCAG 662

249 Val.....ThrPheAspAsnProAlaPheGluProGlyAspThrSe 262

663 TGGCAGCGACATTCATTCATGACCGCGCGCGCAATTAACCGTGT 712

262 rLeuThr.....PheGluProAla.....AspThrAlaP 272

713 GGACCATCAATATCAAGATGTAATTAACATTCGGCGTTCGTTCCACA 762

272 roAspProAspPheLeuAspIleValArgLeuHisArgProAlaLeuThr 288

763 GCGCGTGTACACACGCGCGCGTGTGATTCGCTAGGTTCCTCAAGTCAA 812

289 SerArgArgGlyThrValArgPheSerArgLeuGlyLysAlaIleThr 305

813 CAAACCGCGCTCTGCTACCGTTCGTTGGTGGCGAAGTA..... 852

305 tPheThrArg...ArgGlyThrGlnIleGlyAlaGlnValHisIleThr 321

853 .....TGCAAATTCAGCGCGCGCAATTCGTTGATGACACACACCGC 894

321 isAspIleSerAsnIleThrAlaThrGluAspIleGlnMetGlnProLeu 337

895 GTGATTTCCGCTCGGTATTTGAAGCGCGGATTTACAAAGCGCGCACA 944

338 LeuThrSerLeuSer.....ThrAspGlyLeuThrAsn 348

945 TTAATTTGG...GACGCTACCAATCAATTCGCTTATTCGAAAGAAAGGCC 991

348 P.IleTyAlaAspAlaAspIleAspAsnAlaMetLeuHisThrThrSer 364

992 GCAGCAAGAGAGCTGTCGCGTGGTTCGCGCGCGCGCAAAATCTCC 1041

365 HisThrGlySerThrGlyPro.....ArgSerHisLeuSerPheP 378

1042 ATCAGCGGTACCAACCTCGCGCATTTCTGAAAACAAATCTTCAAGT 1091

378 oSer.....IleProSerThrValSer.....ThrIleTySerAsn 391

1092 CAACACAGCGGTCAAGCGCGCG.....ACGCGCGCA 1123

391 hrThrIleProPheThrThrSerThrAspIleProValThrThrGlyPro 407

1124 TGTGCGGATTCGATTTACAGCGCGCGTGAATGCCCTTGATATCTGCC 1173

408 AspIleValIleuProThrAlaSerProAsnLeuProPheVal.....Pr 422

1174 ACCC 1177

1111





```

236 ethGlyMetLysAsnThrAspIleuHrHisLysMetArgThrG 253
836 TTTTG.....GTCGCAAAATATCGCAAT... 861
253 IuLeuLysAspHisAlaAlaAlaSerGlyIleLysLeuHrYrLeuAla 269
862 .....ACTGGGGCAATTGGTTGACACAGACAGACCCGGT 896
270 PheIleIleLysAlaValAlaLysSerLeuArgAspMetProAsnIleAs 286
897 GATTTCGGTCGTTGATTAACGGCCGAT..... 927
286 nValArgGlyAspPheAlaAsnAsnLysIleGlnPheMetHisAsnIleA 303
928 .....ACA 930
303 snIleGlyIleAlaValAspThrProAsnGlyLeuMetValProValIle 319
931 CAAGCGCGCAGCATTTATTTGGACGCTACACAAATCAGATTTCCGTTAT 980
320 LysGlyAla...AspHisLeuSerValPheGluIleAlaIleLysIleSe 335
981 CGAAGAAGCGCGCAGCAAGAGCTGTTCGGCTGGCGCCGCGACCGCG 1030
335 rglu.....LeuAlaAsnLysAlaL 342
1031 ACAAAATATCCATACACGCTACA.....ACCCTC 1059
342 yAspGlyLysLeuHrArgAlaGluMetThrGluAlaLThrPheHrVal 358
1060 GGCATTTCTGTAACAAACAACTCTTCAAGTTCAACACAGCCGTCACAGG 1109
359 SerAspPheGlySerValGlyLeuAspYrAlaHrProIleIleAsnSe 375
1110 CGCGACCGCGCCGATGTCGCTGACTTACGAGCG..... 1149
375 rProGluSerAlaIleLeuGlyValGlyHrMetSerGlnHrProLeuT 392
1150 .....GTGATCCCTTGATATC 1167
392 YrIleAsnGlyLeuGluGlnLysArgPheIleMetProLeuSerMet 407
seq_name: SwissProt_40:FA5L_SCHPO
seq_documentation_block:
ID FA5L_SCHPO STANDARD; PRT; 2073 AA.
AC Q90UG0; P78799;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fatty acid synthase subunit beta (EC 2.3.1.86) [Includes: 3-
DE hydroxypalmitoyl-[acyl-carrier-protein] dehydratase (EC 4.2.1.61);
DE Enoyl-[acyl-carrier-protein] reductase (EC 1.3.1.9); [ACYL-
DE carrier-protein] acetyltransferase (EC 2.3.1.38); [ACYL-carrier-
DE protein] malonyltransferase (EC 2.3.1.39); S-acyl fatty acid synthase
DE thioesterase (EC 3.1.2.14)].
GN FA5L OR SPAC926.09c.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., AND SUBUNIT.
RX MEDLINE=98359895; PubMed=9593066;
RA Niwa H., Katayama E., Yanagida M., Morikawa K.;
RT Cloning of the fatty acid synthetase beta subunit from fission yeast,
RT coexpression with the alpha subunit, and purification of the intact,
RT multifunctional enzyme complex".
RL Protein Expr. Purif. 13:403-413(1998).
RP SEQUENCE FROM N.A.
RC STRAIN=972;

```

```

RA Wood V., Rajandream M.A., Barrell B.G., Ramsperger U., Pohl T.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1725-2073 FROM N.A.
RC STRAIN=PR745;
RX MEDLINE=98162722; PubMed=9501991;
RA Yoshioke S., Kato K., Nakai K., Okayama H., Nojima H.;
RT "Identification of open reading frames in Schizosaccharomyces pombe
RT CDNA.";
RL DNA Res. 4:363-369(1997).
CC -1- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
CC LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
CC THE BETA SUBUNIT CONTAINS DOMAINS FOR: [ACYL-CARRIER PROTEIN]
CC ACETYLTRANSFERASE AND MALONYLTRANSFERASE, S-ACYL FATTY ACID
CC SYNTHASE THIOESTERASE, ENOYL-[ACYL-CARRIER PROTEIN] REDUCTASE, AND
CC 3-HYDROXYPALMITOYL-[ACYL-CARRIER PROTEIN] DEHYDRATASE.
CC CATALYTIC ACTIVITY: ACETYL-COA + N MALONYL-COA + 2N NADPH =
CC -1- LONG-CHAIN FATTY ACID + (N+1) COA + N CO(2) + 2N NADP(+) +
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
CC acetyl-[acyl-carrier protein].
CC -1- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC malonyl-[acyl-carrier protein].
CC -1- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
CC = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
CC -1- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NAD(+) = trans-
CC 2,3-dehydroacyl-[acyl-carrier protein] + NADH.
CC -1- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
CC carrier protein] + oleate.
CC -1- SUBUNIT: FATTY ACID SYNTHETASE ARE [ALPHA(6)BETA(6)] HEXAMERS OF
CC TWO MULTIFUNCTIONAL SUBUNITS (ALPHA & BETA).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AB010274; BA36384.1; -
DR EMBL: AB110469; CAB54157.1; -
DR EMBL: D89148; BA13810.1; -
DR InterPro: IPR001227; Acyltransf domain.
DR InterPro: IPR003965; Fatty acid synth.
DR InterPro: IPR002539; Maoc dehydratase.
DR Pfam: PR00698; Acyl_transf. 1.
DR Pfam: PR01575; Maoc-dehydratase. 1.
DR PRINTS: PR01483; FASYNTHASE.
KW Fatty acid biosynthesis; Multifunctional enzyme; Oxidoreductase;
KW Transferrase; Hydrolase; Lyase; NAD; NADP.
FT DOMAIN 1 459 ACETYL TRANSFERASE.
FT DOMAIN 470 858 ENOYL REDUCTASE.
FT DOMAIN 1155 1644 DEHYDRATASE.
FT DOMAIN 1645 2073 MALONYL/PALMITOYL TRANSFERASE.
FT ACT_SITE 270 270 ACETYL TRANSFERASE (BY SIMILARITY).
FT ACT_SITE 1361 1361 DEHYDRATASE (POTENTIAL).
FT ACT_SITE 1828 1828 MALONYLTRANSFERASE (BY SIMILARITY).
FT CONFLICT 222 P -> R (IN REF. 1).
SQ SEQUENCE 2073 AA; 23055 MW; D92827083BE7C7C4 CRC64;

```

## alignment\_scores:

```

Quality: 110.50 Length: 465
Ratio: 0.521 Gaps: 25
Percent Similarity: 45.591 Percent Identity: 20.215

```

## alignment\_block:

```

US-09-303-518d-125 x FA5L_SCHPO
Align seg 1/1 to: FA5L_SCHPO from: 1 to: 2073

```

```

217 TCAGCGAAATGCGCCGCGATTCACGCTGGCGAAAGCGGCTACTTCAGTC 266
:::|||||::: :::::||||| :::::|||||

```

```

948 ThrGlyLysIeu.....MetArgArgIleGluGluArgPheThrGlnAs 962
267 AGTC.....GTGATTCGGCTTGAGAGCAGCAGCAATGAG..... 303
962 pValGlyLysThrThrLeuIleGluAsnPheGluAspLeuAsnProT 979
304 .....TTTGAACGCTACGACGACCTGAAGCGCTGGCA 333
979 yrProValAlaAlaArgPheLeuAsnPalatyrProGluAlaSerThrGln 995
334 AACTTAAGCGCGGAGAGAGTGGCCGCAACCTGATCCAAATCCGGTTGTG 383
996 AspLeuAsnThrGlnAspAlaGln.....PheP 1005
384 GACGCGCTGGCGACCGCTCGCTCAGCAAA.....ATTCTG 421
1005 eTyrSerLeuCysSerAsnProPheGlnLysProValProPheIleProA 1022
422 CCGTCGAT..... 429
1022 LaIleAspAspThrPheGluPheThrPheLysAspSerLeuTrpGln 1038
430 GCCGAGCCGCTGCCCATCTTCGTCAT.....GGAT 461
1039 SerGluAspLeuAlaAlaValAlaGlyLysAspValGlyArgValAlaI 1055
462 GGACACCAATCCGCTGGCTGCCGACCTAGCGCATTCATAAAGACCG 511
1055 eLeuGlnGlyProMetAlaAlaLysHisSerThrLysValAsnGluProA 1072
512 CCGAGGATTTCAAAACGCGCGCTGTGTATAGACCGCTTGACGAGCAAGC 561
1072 LaLysGlu.....LeuLeuAspGlyIleAsnGluThr 1082
562 AAAATC...CATGTTTGAAGCA.....GCTGGCCGAGAC.....GT 596
1083 HisIleGlnHisPheIleLysLysPheThrLysIleAspGluLysIle 1099
597 GCCGCTGAAATGCTGCCACATCGAACAATCGCGCGCGCGC 646
1099 eProIle.....ValGluThrPheLysGlyValP 1109
647 ATCTGCGCGTTTGAAT..... 663
1109 roProValAsnValSerHisLysSerLeuGlnSerValThrGlu 1125
664 .....GGCAGCACATTCATTCATGAGCGCGCGCGGAGATAAAC 707
1126 GluAlaGlySerLysValLysLeuProGluIleGlySerAsnSerAl 1142
708 CGTG..... 711
1142 AleuProSerLysLysLeuTrpPheGluLeuLeuAlaGlyProGluTyrT 1159
712 .....TGAC 717
1159 hrTrpPheArgAlaIlePheThrThrGlnArgValAlaLysGlyTyrLys 1175
718 ATCAATTTATCAAGATGTAATTCATTCGCGCTTTTGCAACGCGCG 767
1176 LeuGluHisAsnProVal.....ArgArgIlePheAlaPro..... 1187
768 TCTGAACACGCGCGCTGATTCGCTTGTGTTCTCAAGTCAACAAC 817
1188 ArgTyrGlyGlnArgAlaValValLysGlyLysAspAsnThrVal 1204
818 CCGCGCTCTTGGTAC.....GTTTGGGTGCAAAATA 852
1204 alGluLeuTyrGluThrGlnSerGlyAsnTyrValLeuAlaAlaArgLeu 1220
853 TGGCAAAATTAACGCGCGCAATTTGTTGACACAGACCGGTCATTTTC 902
1221 Ser.....TyrAspGlyGluThrIleValValSe 1230

```

```

903 CGGTTGGTATTGAACGGCGGATTCACAGCGCGCAC...GATTATT 949
1230 rMetPheGluAsnArgAsnAlaLeuLysGlyGluValHisLeuAspPheI 1247
950 TGGGACGCTAC.....CACAAATCAGATTTCGGTTATCGCA 984
1247 euPheLysTyrGluProSerAlaGlyTyrSerProValSerGluIleLeu 1263
985 GAAGCGCGGACCAAGAGCTG.....TTGCGCTGGCTT..... 1017
1264 AspGlyArgAsnAspArgIleLysHisPheThrPalaleuTrpPheG 1280
1018 ....GCGCGCAGCGGACAAATTAATCCATACAGCGGTTCACAACTCGGCC 1063
1280 yGluGluProTyrProGluAsnAlaSerIleThrAspThrPheThrGlyP 1297
1064 ATTTCCTGAAA.....AACAAACTTCATCAATGCACACAGCCGTC 1104
1297 roGluValThrValThrGlyAsnMetIleGluAspPheCysArgThrVal 1313
1105 AACGCGCGGACCGCGCGCATGTGCGCATTTGCTTACTTACGACGCGGTAT 1154
1314 GlyAsnHisAsnGluAlaLysThrLysArgAlaIleArgLysArgMetAl 1330
1155 GCCCTTGGAT.....A 1165
1330 aProMetAspPheAlaIleValAlaGlyTyrPheAlaIleThrLysAlaI 1347
1166 TCGTGGCCGACCGCTGTTGGCGGATTTA.....ATGCGCGCATATCC 1209
1347 lePheProLysAlaIleAspGlyAspLeuLeuArgLeuValHisLeuSer 1363
1210 GACAGCGCGGACGATTTGGTCTTGAATTCGACGAGAGAGAC 1254
1364 AsnSerPheArgMetValGlySerHisSerLeuMetGluGlyAsp 1378
seq_name: SwissProt_40:CHLN_SYNP7
seq_documentation_block:
ID CHLN_SYNP7 STANDARD; PRT; 466 AA.
AC P54208;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Light-independent protochlorophyllide reductase subunit N
DE (EC 1.18.-.-) (Lr-POR subunit N) (DPOC subunit N).
OS CHLN OR FRXC.
OS Synchococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OC Bacteria; Cyanobacteria; Chroococcales; Synchococcus.
OX NCBI_TaxID=1140;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95388763; PubMed=7659748;
RA Ronen-Tarazi M., Lieberman-Hurwitz J., Gabay C., Orus M.I., Kaplan A.;
RT "The genomic region of rbcLs in Synchococcus sp. PCC 7942 contains
RT genes involved in the ability to grow under low CO2 concentration and
RT in chlorophyll biosynthesis."
RL Plant Physiol. 108:1461-1469(1995).
CC -!- FUNCTION: Uses Mg-ATP and reduced ferredoxin to reduce ring D of
CC protochlorophyllide (Pchl) to form chlorophyllide a (Chlide)
CC (By similarity). This reaction is light-independent.
CC -!- PATHWAY: Light-independent chlorophyll biosynthesis.
CC -!- SUBUNIT: Protochlorophyllide reductase is thought to be composed
CC of three subunits: chlL, chlN and chlB. Could form a
CC heterotrimer of two chlB and two chlN subunits (By similarity).
CC -!- SIMILARITY: BELONGS TO THE BCCH / CHLN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

[illegible]

CC This SWISS-PROT

CARBONYL ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-glucose residues successively from non-reducing ends of the chains with release of beta-D-glucose.  
 - SIMILARITY: TO: S.POMBE SPBC215.13.  
 - SIMILARITY: SOME, TO S.POMBE SPCC285.13C.  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation.

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----  
 DR EMBL; Z38061; CA86176.1; -  
 DR EMBL; M16164; AAA35014.1; -  
 DR EMBL; M16165; AAA35015.1; -  
 DR EMBL; X13857; CAA32069.1; -  
 DR PIR; B26877; B26877.  
 DR PIR; A26877; A26877.  
 DR PIR; S48478; S48478.  
 DR SGD; S0001458; M0C1.  
 KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;  
 KW Signal; Multigene family.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 1367 GLUCOAMYLASE SL/SZ.  
 FT DOMAIN 210 1367 SER/THR-RICH.  
 FT CARBOHYD 817 817 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 874 874 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1367 AA; 136110 MW; 91C0E2DBD61AA9D CRC64;

alignment\_scores:  
 Quality: 107.00 Length: 545  
 Ratio: 0.476 Gaps: 18  
 Percent Similarity: 41.284 Percent Identity: 19.633

alignment\_block:

US-09-303-518D-125/rev x AMYH\_YEAST ..

Align seg 1/1 to: AMYH\_YEAST from: 1 to: 1367

1337 TCCTTCATAGTGTTCACGACTTTCGCAACAGCGGCGGATTCGTA 1288  
 |||||  
 891 SerheserThrGlyThrValThrProserSerSerlysty.... 905  
 1287 TTTGCCGGGCGAGACGAGCGCAACGAGGCTTCCTCGCAAT 1238  
 |||||  
 906 ....ProGlySerThrGlyThrSerValThrSerThrThrGluThr 921  
 1237 CCAAGCAACCAATGCTGCGCTGCGTA.....TCGCG 1200  
 |||||  
 921 hrIleValProThrLysThrThrSerValThrThrProserThrThr 937  
 1199 ACGATTAATCGCGCAAGAGGCGGCGGATTCGCAAGGCGATCAC 1150  
 |||||  
 938 ThrIleThrThrValCysSerThrGlyThrAsnSerAlaGlyGluThr 954  
 1149 GCGGCTGTAAGTACCAATCGGCGGCGGCGGCGGCGGCTTGCAGG 1100  
 |||||  
 954 rThSerThrCysSerProLysThrValThrThrValProthrThr 971  
 1099 CTGCTGTAACCTGAAGATTGTTTTCAGGAAATGCGCGAGGCTTGA 1050  
 |||||  
 971 hrThrThrSerValThrThr.....SerSerThrThrThrIle 983  
 1049 CGCGGATGAGGATTTTCGCGGCGGCGGCGGCGGCGGCGGAGCTC 1000  
 |||||  
 984 ThrThrThrValCysSerThrGlyThrAsnSerAlaGlyGluThrSe 1000  
 999 TTTGCTGCGGCGCTTCGATACGGAATCGATTGTGTAGCGTCCA 950  
 |||||  
 1000 rGlyLysSerProLysThrIleThr..... 1008  
 949 AATTAATCGGCGGCTGTGTAATCGGCGGCTTCATACCGAACGGA 900  
 |||||  
 1009 ....ThrThrValProCysSerThrSerProserGluThrAlaSerGlu 1023  
 900 ..... 900  
 1024 SerThrThrThrSerProThrThrProValThrValThrValSerThr 1040

899 .....ATCAGCGGCTGTCTGTCACCAACCAATGCGCGGAGTATT 857  
 |||||  
 1040 rValValThrThrGlyThrSerThrThrLysProGlyGluThr 1057  
 |||||  
 856 GCGATCTTCGCGACCAACAGGTAACGAGCGCGGCTGTGACT 807  
 |||||  
 1057 hrThrThrThrPheValThrLysAsnIle...ProThrThrThrLeuThrThr 1072  
 |||||  
 806 TGAGAACCACTTACGCAATCAGCGCGCTGCTGTCAGAGCGGCTTGC 757  
 |||||  
 1073 IleAlaProThrProSerValThr.....ThrValThr 1083  
 756 AACAACAGCGCAATGTAATTCATCTGATATGATGATGTCACACG 707  
 |||||  
 1083 rAsnPheThrProThrThrThrIleThrThr.....ThrV 1094  
 706 TTTTATTCGCGCGACCGGCTCGATGAATGATGCTGCTCCACAA 657  
 |||||  
 1094 aLysSerThrGlyThrAsnSer.....AlaGlyGluThr 1105  
 656 CCGGAGGATGCGGCGG..... 639  
 1106 ThrSerGlyCysSerProLysThrValThrThrValProCysSerThr 1122  
 639 ..... 639  
 1122 rGlyThrGlyGluThrThrThrGluAlaThrThrLeuValThrThrAlaV 1139  
 639 ..... 639  
 1139 aThrThrThrValValThrThrGluSerSerThrGlyThrAsnSerAla 1155  
 638 .....CGAATTCATGTGT 625  
 1156 GlyLysThrThrThrGlyThrThrThrLysSerValProThrThrThrY 1172  
 624 TTCGATGTTGCGACATTTTCAGACGCGACGTCGCGCAGCTGCTTAC 575  
 |||||  
 1172 ThrThrLeuAlaProSerAlaProValThrProAlaThrAsnAlaValP 1189  
 574 AACAATGATTTTGGCTGCGTCAACAGGCTCAATACCAACAGCGCGGT 525  
 |||||  
 1189 roThrThrThrIleThrThrThrGlyCysSerAlaAlaThrAsnAlaIleGly 1205  
 524 TTGAATCCTCGCGGCGCTTTCATATGACCGTAGGCTGCGGCGCAG 475  
 |||||  
 1206 GluThrThrSerValCysSerAlaLysThrIleValSerSerAlaSerAl 1222  
 474 CGGATTTGGTCCATCGCATTCGCAAGATGCGCAACGCGCTGCGCATGA 425  
 |||||  
 1222 agly.....GluAsnThrAlaProS 1229  
 424 CGGCGAATTTTGTGTAACGAGCGGCGGCGGCGGCGGCGGCGGAGCTC 375  
 |||||  
 1229 erAlaThrThrProValThrThrAlaIleProThrThrVal..... 1242  
 374 GATTGATCAGGTTGCGGCGGCGGCGGCTTTCGCGGCTTAATTTGCCAGCGC 325  
 |||||  
 1243 .....IleThrThrThrThrSerValGlyThrAsnSerAl 1254  
 324 TTCAGGCTGCTAGCTTCGAATCGATTCGTCGTCGCTTCGCTTCACGCGCA 275  
 |||||  
 1254 aglyGluThrThrThrGlyThrThrThrLysSerIleProThrThrThrI 1271  
 274 TCACGACTGACGTAAGTACGCGCTTTTCGCCACGCGTAATTCGCGCGAT 225  
 |||||  
 1271 leThrThr.....Leu 1274  
 224 TTGCGCT.....GAAGCGGCGGCGGTAACAC 199  
 |||||  
 1275 IleProGlySerAsnGlyAlaLysAsnThrGlyThrValAlaThrAlaThr 1291



```

198 CACGCC.....GGATCTTTTGTCTT 176
1291 rnsnProIleSerIleThrThrSerGlnLeuAlaThrAlaSerA 1308
175 CAACAGCACTGGCCTTTTGGAGCATGGCTTCTTCTTACCTTTCATC 126
1308 laseSerValAlaProValAlaThr...SerProSerLeuThrGly... 1322
125 GAGGGCGCATATTCCTCGCCAAAGCAAGCAAGCTTGGTAT 76
1323 .....ProLeuGlnSerAlaSerGlySerAlaValAlaThrTy 1335
75 GGCCGGCCGTGTAAACGGCT.....TGCTCGGCTGTGCCG 38
1335 rSerValProSerIleSerSerThrTyGlnGlyAlaAlaAlaSerIleLysV 1352
37 CGATGGCAGGTTTAGACCTTTTGGATTATTC 3
1352 allLeuGlyAsnPhenMetTrpLeuLeuAlaLeu 1363

```

seq\_name: SwissProt\_40:PPSA\_PYRFU

seq\_documentation\_block:

ID\_PPSA\_PYRFU STANDARD; PRT; 817 AA.

AC P42850; Q59672;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Probable phosphoenolpyruvate synthase (EC 2.7.9.2) (Pyruvate, water dikinase) (PEP synthase).  
 GN PPSA OR PF0043.  
 OS Pyrococcus furiosus.  
 OC Archaea: Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
 NX NCBI\_TaxID=2261;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=Vc1 / DMS 3638;  
 RX MEDLINE=95129854; PubMed=7828869;  
 RA Robinson K.A., Schreier H.J.;  
 RT "Isolation, sequence and characterization of the maltose-regulated mltA gene from the hyperthermophilic archaeum Pyrococcus furiosus.";  
 RL Gene 151:173-176(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Vc1 / DMS 3638;  
 RC MEDLINE=95354939; PubMed=7628701;  
 RA Jones C.B., Fleming T.M., Piper P.W., Littlechild J.A., Cowan D.A.;  
 RT "Cloning and sequencing of a gene from the archaeon Pyrococcus furiosus with high homology to a gene encoding phosphoenolpyruvate synthetase from Escherichia coli.";  
 RL Gene 160:101-103(1995).  
 CC -1- CATALYTIC ACTIVITY: ATP + pyruvate + H(2)O = AMP + 4-phosphoenolpyruvate + phosphate.  
 CC -1- PATHWAY: ESSENTIAL STEP IN GLUCONEOGENESIS WHEN PYRUVATE AND LACTATE ARE USED AS A CARBON SOURCE.  
 CC -1- SIMILARITY: BELONGS TO THE PEP-UTILIZING ENZYMES FAMILY.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

CC EMBL: U08376; AAA81512.1; -;  
 CC EMBL: X80819; CAA56785.1; -;  
 CC HSSP: P22983; IDIK.  
 CC InterPro: IPR000121; PEP\_utilizers.  
 CC InterPro: IPR002192; PPK\_N\_term.  
 CC Pfam: PF00391; PEP\_utilizers\_1.  
 CC Pfam: PF02896; PEP\_utilizers\_C\_1.  
 CC Pfam: PF01326; PPK\_N\_term; 1.

DR ProDom: PD000940; PEP\_utilizers; 1.  
 DR PROSITE: PS00370; PEP\_UTILIZERS\_PHOS\_SITE; 1.  
 DR PROSITE: PS00742; PEP\_ENZYMES\_2; 1  
 KW Transferase; kinase; ATP-binding; phosphorylation.  
 FT MOD\_RES 442 442 PHOSPHORYLATION (BY SIMILARITY).  
 FT DOMAIN 809 815 POLY-GLO.  
 FT CONFLICT 747 747 Q -> K (IN REF. 2).  
 SO SEQUENCE 817 AA; 90485 MW; 582694CERF13CT4BA CRC64;

#### alignment\_scores:

| Quality:            | 105.50                           | Length:           | 468    |
|---------------------|----------------------------------|-------------------|--------|
| Ratio:              | 0.451 <td>Gaps:</td> <td>22</td> | Gaps:             | 22     |
| Percent Similarity: | 50.000                           | Percent Identity: | 21.368 |

#### alignment\_block:

US-09-303-518D-125 x PPSA\_PYRFU ..

Align seg 1/1 to: PPSA\_PYRFU from: 1 to: 817

```

67 GGCCCGCCATTCACGAGTCGCGTTCGTCGGAAGATATCCGCTAT 116
117 GCGCCCGCTCGATGAAAGTCAGAGAGCGCATGCCGCAAAAAGGCCAG 166
252 lGluGlnAlaValAlaSerGlyAlaValThrProAspGluTrpIleValG1 268
117 GCGCCCGCTCGATGAAAGTCAGAGAGCGCATGCCGCAAAAAGGCCAG 166
268 uLysGlyThrTrpLysIleLysGluValIleAlaLysLysGluValM 285
167 TGCTGTTTGAAGACAAAAGATCCG.....GGCGTGTGTTT 204
285 etValIle.....ArgAsnProGluThrGlyArgGlyThrValMet 298
205 ACTGCCCGCGCTTCAGCAAAATCCGCGCGATTCCACCGGC..... 246
299 Val.....LysValAlaGlu...TyrLeuGlyProGluTr 309
247 .....GAAAGCGGCTACTT.....CAGTCAGTCGATTCGCGTGAG 286
309 pValGlnLysGlnValLeuThrAspGluGlnIleIleGluValAlaLysM 326
287 GCAAGCAGCAATCGAGTTTGAACGCTACGCACTGACCGATGACGCA... 333
326 etGlyLysIleGluAspHisTyrGlyTyrProGlnAspIleGluTrp 342
334 AACTTAAGCGCGCAGAGAGTGCAGCCGCAACCTGATCAATCCGTTTG 383
343 AlaTyrAspLysAspAspGlyLysLeuTyrIleValGlnSerArgPro 359
384 GACTGGCGTCGCGACCGCGCTTCAGCAAAATTCCTGCGCGTGCATGCC 433
359 eThrThrLeuLysGluGlnAlaThrAlaGlu...GluAlaGluGluValG 375
434 AGCGTTCGCAATTCGTCATGATGATGATGATGATGATGATGATGATGAT 480
375 lGluGlnAlaValIleLeuLysGlyLeuGlyAlaSerProGlyIleGly 391
481 GCGGACCCCTACGCTCATTTCAAGAAGACCGCCGAG.....GATTCAA 524
392 lAlaGlyArgValAlaValIlePheAspAlaSerGluIleAspLysVal 408
525 ACCGCGC...CTGTGTATTTAGCCGTTTGACCGCAAGCAAAATCCAG 571
408 sGluGlyAspIleLeuValThrThrThrAsnProAspMetValProA 425
572 TTTGTAGGACAGCTGGCGCAGAGATGCGCTGCAAAATCT..... 612
425 lMetLysArgAlaAlaAlaIleValThrAspGluLysGlyArgThrSer 441
613 ...GCCAATTCGAACACATGATTCGCGCGCGCCGCACTGCGCGTTT 659
442 HisAlaAlaIleValSerArgGluLeuGlyIlePro..... 453
660 GAGTGCAGCAGACATTCATTCATGACGCGGTCGCGCGGCAATAACCG 709

```

```

453 ..... 453
710 TGTGACCATCAATTATCAAGATGTAATACATTGCGCGTTTGTTCGA 759
454 ..... CysValValGlyThrLysGluLeu 461
760 ACAGCGCGCTGACACGAGCGCGTGAATGCCCTAGTGGTTCACA.. 807
462 ThrLysLysLeuLysThrGlyMetLysValThrValAspGlyThrArgI 478
808 ...GTCACAACACCGCGCGCTTTCGCTACCGCTTTGGGCGGAAGAT 853
478 yLeuValLysGlyLysLeuLysSerLeuValLysLysGluGlu 495
854 CCCAATTACTGCGCGCATGTTGACACGACACCGCGGTGTTTC 903
495 lAlaLysAlaGlyLysGlyLysValValAlaGlyAlaProLeuValThr 511
904 GGTTCGCTATG..... 915
512 GlyThrMetValLysValAsnValSerMetProGluValAlaGluArgAl 528
916 .....AACGCGCG.....ATTACACACGCGCGCAGATT 946
528 aAlaAlaThrGlyAlaAspGlyValGlyLeuLeuArgAlaGluHisMetI 545
947 ATTTCGAGCGCTACCAATCAGATTTCGCTTACGACACGCGCGCAGC 996
545 lLeuSerLeuGlyLysProLysPheLysGlyLysGlu 561
997 AAGAGCTGTTGCGCTGGTGGCGCGCAGCGCAACAACTCCATCCATCAC 1046
562 GluGluLeuValGluLysLeuAlaGlyLysGlyLysValAlaAla.. 577
1047 GGTTCACACCTTCGCGCTTCTGAAAAACAACCTTTCACAGTTCAACA 1096
578 AlaPheLysProArgProVal..TrpTyrArgThrLeuAspAlaProTh 593
1097 CAGCGCTCAACGCGCGCGCGCGCATTCGATGAGTACTGATCAGAG 1146
593 rAsnGluPheArgGluMetPro..GlyGlyLysPheLysValProGluGlu 609
1147 CGCGTGATCCCTTGATATCTCGCC.....CACCT 1178
609 rGAsnProMetLeuGlyTrpArgGlyLeuArgArgLysLeuAspLysPro 625
1179 GCTTTGGCGCGATTATCGTGGCGATACCGACGCGCGCATTCG 1228
626 GluLeuLeuArgAlaGluPheLysAlaLeuLysValValGluLysGlu 642
1229 GTTGGTGAATTGGACGAGAACACCGCTTGTGACAGCTGCTG 1277
642 yTyrAsnAsnLysGly.....ValMetLeuProLeu 653
1278 .....CCGCGCAATACGATACGCGCGCTGTTGCGCAAGTCT 1319
653 aLSerHisProGluGlnLeuArgLysAlaLysArgLysAlaArgLysGluVal 669
1320 GGA 1322
670 Gly 670

```

seq\_name: Swissprot\_40:SON\_MOUSE

seq\_documentation\_block:

ID SON\_MOUSE STANDARD; PRT; 2404 AA.  
AC 090X47; 090X45; 090X46; 090X47;  
DT 01-MAR-2002 (Rel. 41, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE SON protein.  
GN SON.

```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN
[1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC STRAIN=129/Sv;
RC MEDLINE=20408886; PubMed=10950926;
RA Wynn S.L., Fisher R.A., Pagel C., Price M., Liu Q.Y., Khan I.M.,
RA Zammit P., Dadrach K., Mazrani W., Kessling A., Lee J.S., Bijuvel L.;
RT "Organization and conservation of the GART/SON/DONSON locus in mouse
RT and human genomes.";
RL Genomics 68:57-62(2000).
[2]
RP SEQUENCE OF 1-116 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Hippocampus, Small intestine, and Tongue;
RC MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Felschmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker G., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC
CC -I- FUNCTION: Transcriptional repressor. Binds to the consensus DNA
CC sequence: 5'-GAGCTA[CG]AGCC-3'. Might protect cells from
CC apoptosis. Might be involved in pre-mRNA splicing (by similarity).
CC -I- SUBCELLULAR LOCATION: Nuclear (by similarity).
CC -I- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are
CC produced by alternative splicing.
CC -I- TISSUE SPECIFICITY: Widely expressed.
CC -I- DOMAIN: Contains 8 types of repeats which are distributed in 3
CC regions.
CC -I- SIMILARITY: CONTAINS 1 G-PATCH DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 DBM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC
DR EMBL, AF193606; AAF23120.1; -
DR EMBL, AF193595; AAF23120.1; JOINED.
DR EMBL, AF193596; AAF23120.1; JOINED.
DR EMBL, AF193597; AAF23120.1; JOINED.
DR EMBL, AF193598; AAF23120.1; JOINED.
DR EMBL, AF193599; AAF23120.1; JOINED.
DR EMBL, AF193600; AAF23120.1; JOINED.
DR EMBL, AF193601; AAF23120.1; JOINED.
DR EMBL, AF193602; AAF23120.1; JOINED.
DR EMBL, AF193603; AAF23120.1; JOINED.
DR EMBL, AF193604; AAF23120.1; JOINED.
DR EMBL, AF193605; AAF23120.1; JOINED.
DR EMBL, AF193607; AAF23121.1; JOINED.
DR EMBL, AK019312; BAB31658.1; -
DR EMBL, AK019081; BAB31536.1; -
DR EMBL, AK008478; BAB25691.1; -

```

|    |  |
|----|--|
| DR | EMBL:AK008256; BAB25562.1; -.                      |
| DR | MGD; MGI:98353; Son.                               |
| DR | InterPro: IPR001159; DS_RBD.                       |
| DR | InterPro: IPR000467; G_patch.                      |
| DR | Pfam: PF00035; dsrm; 1.                            |
| DR | Pfam: PF01585; G_patch; 1.                         |
| DR | SMART: SM00443; G_patch; 1.                        |
| DR | PROSITE: PSS0137; DS_RBD; 1.                       |
| DR | PROSITE: PSS0174; G_PATCH; 1.                      |
| KW | RNA-binding; DNA-binding; Nuclear protein; Repeat; |
| KW | Alternative splicing.                              |
| FT | DOMAIN 721 850                                     |
| FT | 687 943  |
| FT | DOMAIN 961 1080                                    |
| FT | REPEAT 961 966                                     |
| FT | REPEAT 969 974                                     |
| FT | REPEAT 976 981                                     |
| FT | REPEAT 985 990                                     |
| FT | REPEAT 993 998                                     |
| FT | REPEAT 1001 1006                                   |
| FT | REPEAT 1010 1015                                   |
| FT | REPEAT 1018 1023                                   |
| FT | REPEAT 1026 1031                                   |
| FT | REPEAT 1035 1040                                   |
| FT | REPEAT 1044 1049                                   |
| FT | REPEAT 1055 1060                                   |
| FT | REPEAT 1066 1071                                   |
| FT | REPEAT 1075 1080                                   |
| FT | DOMAIN 1101 1133                                   |
| FT | 1910 1979  |
| FT | REPEAT 1910 1916                                   |
| FT | REPEAT 1938 1944                                   |
| FT | REPEAT 1945 1951                                   |
| FT | REPEAT 1952 1958                                   |
| FT | REPEAT 1959 1965                                   |
| FT | REPEAT 1966 1972                                   |
| FT | REPEAT 1973 1979                                   |
| FT | DOMAIN 1919 1990                                   |
| FT | REPEAT 1919 1937                                   |
| FT | REPEAT 1980 1990                                   |
| FT | DOMAIN 1991 2017                                   |
| FT | 2283 2329  |
| FT | DOMAIN 2349 2404                                   |
| FT | VASAPLIC 2086 2086                                 |
| FT | VASAPLIC 2087 2404                                 |
| SO | 2404 AA; 261428 MW; 44BBE28BD3FC01D9 CR664;        |

```

Alignment_scores:
      Quality: 105.50      Length: 506
      Ratio: 0.444      Gaps: 24
Percent Similarity: 48.024      Percent Identity: 23.123

Alignment_block:
US-09-303-518D-125 x SON_MOUSE ..

Align seg 1/1 to: SON_MOUSE from: 1 to: 2404

18 AGGCTTAAACCTGCGCCATCGCGGGACGACGCAAGC.....CG 58
|||||..... |||
1860 ArgSerArgSerIysSerIArgIArgSerValSerLysGluLysArg 1876
|||||..... |||

59 TTTCAGAGGCGCGCCGATACCGAAGTGCCTGTGGCGAGAAATAT 108
||| ||| ||| ||| |||..... |||
1876 GluArgSerIProLysHis...ArgSerIysSerArgIuArgLysArgL 1892
|||||..... |||

109 GCCGGTATGCGCCCTCGATGAAATGCAAGAGAGCGATCCGCTCAAAA 158
||| ... |||..... |||

```

```

894 CGTATTTCCGGTTCGGTATGACGGCGGATTCACAGAGCGCGCAG 943
      ::::::::::::::::::::
2188 nvalpheserleuprosergluval...asp.ilserthr 2203
      ::::::::::::::::::::
944 ATTATTTGGACGCTACCAATCAGATTCGGTATGAGAGAGCGCGC 993
      ::::::::::::::::::::
2204 AlmetSerGluArgAlaLeuAlaGlnArgLeuSerGluAlaAlaPh 2220
      ::::::::::::::::::::
994 AGCAAGAGCTGTTCGGTGGCGCGCGCAGCGGACCAATCTCCAT 1043
      ::::::::::::::::::::
2220 eaapleuglualametermetleuasnargalaglnuargileaspa 2237
      ::::::::::::::::::::
1044 CAGCGGTACACCCGCGCATTTCTGAAACAAACT.....CTCA 1087
      ::::::::::::::::::::
2237 latpalaglnleuanserile.ProgIyGlnPheThrGlySerThrGl 2253
      ::::::::::::::::::::
1088 AGTTCAACACAGCGCGTCAACGGCGGCGCGCATGTGTGC..... 1130
      ::::::::::::::::::::
2253 yvalGlnValleuThrGlnleuGlnleuAlaSnThrGlyAlaGlnAlaT 2270
      ::::::::::::::::::::
1131 .....GATTGGTACTTACGAGCGCGTGATGCGCTTGATATCT 1169
      ::::::::::::::::::::
2270 rpllelylsyAspGlnPheleuArgAlaAlaProvalThrGly..... 2284
      ::::::::::::::::::::
1170 GCCCACCCTGCTTTGGCGCATTTATCGCGCGCATACCGACAGCGCGC 1219
      ::::::::::::::::::::
2285 ...GlyMetGlyAlaValleuMetArgLysMetGlyTyrPargIleuYgl 2300
      ::::::::::::::::::::
1220 AGGCATGGGGTTCCTTGAATGTGACGAGAGAACCTCCCTTGTGACAC 1269
      ::::::::::::::::::::
2300 uGlyleuGlyLysAsnLysGlnLysGlnLysGlnLysGlnLysAlaSP 2317
      ::::::::::::::::::::
1270 TTCCTCTG.....CCCGGCAATACGAAATACGCGCGCTGTGGCA 1313
      ::::::::::::::::::::
2317 helysThrAspArgLysGlyLeuValAlaValGlyGlnArgAlaGlnLys 2333
      ::::::::::::::::::::
1314 AGTGCTGGAAC 1325
2334 ArgSerGlyAsn 2337
seq_name: SwissProt_40: MURF_HAEIN
seq_documentation_block:
ID MURF_HAEIN STANDARD; PRT; 457 AA.
AC P45061;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE UDP-N-acetylmuramoylalanine-D-glutamate-2,6-diaminopimelate--D-alanyl-D-
DE alanyl ligase (EC 6.3.2.15) (UDP-MurNAc-pentapeptide synthetase)
DE (D-alanyl-D-alanine-adding enzyme).
OS MURF OR H11134.
CN Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae.
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleisichmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-L., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fitchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus
RA influenzae Rd."
RL Science 269:496-512(1995).

```

```

CC -1- FUNCTION: INVOLVED IN CELL WALL FORMATION. CATALYSES THE FINAL
CC STEP IN THE SYNTHESIS OF UDP-N-ACETYLMURAMOYL-PENTAPEPTIDE, THE
CC PRECURSOR OF MUREIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanyl-D-
CC glutamyl-meso-2,6-diaminoheptanedioate + D-alanyl-D-alanine = ADP
CC + phosphate + UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-6-
CC carboxy-L-lysyl-D-alanyl-D-alanine.
CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE MURDEF FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: U32793; AAC22789.1; -.
CC DR HSSP: P11880; 1G94.
CC DR TIGR: H11134; -.
CC DR InterPro: IPR000713; Mur_Ligase.
CC DR InterPro: IPR004101; Mur_Ligase.C.
CC DR Pfam: PF01225; Mur_Ligase.1.
CC DR Pfam: PF02875; Mur_Ligase.C; 1.
CC DR Peptidoglycan synthesis; Cell division; Cell wall; Ligase;
CC KW ATP-binding; Complete proteome.
CC FT NP_BIND 109 115 ATP (POTENTIAL).
CC FT SEQUENCE 457 AA; 50073 MW; 095FEF74F27CD20B CRC64;

```

```

alignment_scores:
  Quality: 105.00      Length: 381
  Ratio: 0.550        Gaps: 20
  Percent Similarity: 50.131  Percent Identity: 21.785

```

alignment\_block:

US-09-303-518D-125 x MURF\_HAEIN ..

```

Align seg 1/1 to: MURF_HAEIN from: 1 to: 457
253 CGCGTACTTCAGTCAGTCGATTCGCGTGAAGCGACGACGAATCGA 302
      ::::::::::::::::::::
      11 GlnleuGlnAlaLysleu.....GlyAspGlnAsnValG1 24
303 GTTGAACGCTACGACACCTGAGCGCTGGCAACTTAAGCGGGAAGAG 352
      ::::::::::::::::::::
      24 nvalGlnLysleuAsnThrAspThr..... 32
353 TGGCGCGACCTGATTCATTCGCGTGTGGACGCGCGCGCACCGCT 402
      ::::::::::::::::::::
      33 ....ArgLysSerValSerAsnSerleuPheAlaLeuLysGlyGlu 47
403 CCGTTCACGAAATTCCTGCGCTGCAT.....GCGAGCGCTTGGCAT 446
      ::::::::::::::::::::
      48 LysPheAspAlaHisGlnTyrLeuAspAlaValSerGlnLysAlaLe 64
447 CTGTCGTAATGCGATGACACCAATCCGCTGCTGCCGACCGCTACGCTA 496
      ::::::::::::::::::::
      64 uAlaLeuValAlaGlnGlnGlnAsnSerSerIleSerValProGlnLeu 81
497 TTATCAAGAA.....CGCGCGAGATTTCAACGCGCGCTGTG 537
      ::::::::::::::::::::
      81 aValLysAspThrArgIleAlaLeuGlyGlnLeuAlaLysTyrLeu... 96
538 GTATTGACCGCTTGACCGAAGCAAAATCCATGTTGTGAAGCAGCTGG 587
      ::::::::::::::::::::
      97 ...ArgGlnLysIleAsnProArgThrValAlaMetThrGlySerSerG1 112
588 CGCAACGTCGCGCTGTAATGCTGCCAATGCAACATGAATTCG 637
      ::::::::::::::::::::

```

```

112 ylvsthrValylsgluMetThrAlaSerIleLeuGlnHisThr.... 127
638 GCGGCGCGCAACCTCCGCGGTTCAGTGGCAGCAGATTCATTTCATCGAG 687
128 ..AlaAlaAspSerGluAlaValLeuPheThrAsnGlyAsnPheAsn 143
688 CCGGTGCGCGCAATAAACCCTGTGGACATCAATTATCAAGATTAAT 737
144 AspleGlyVal.....Prole 149
738 TACCATGGCCGCTTTGTTTCAGACAGCCGTGCAACCGAGCGGTGA 787
149 urThrLeuAlaArgLeuThrGluValHisArgPhe.....AlaValI 163
788 TTGGCTAGTGGT.....TTCAGTACACAAACCGCGCTCTTG 828
163 legluLeuGlyAlaAsnHisGlnAsnGluIleAsnIleThrThrIleu 179
829 CGTACCGTTTGGGTGCGAAAGTATGCAAAATTAATGCGCGCGAATG.. 876
180 ValGlnProAsnAlaAlaLeuIleAsnAlaIleAlaProAlaHisLeuG 196
877 .....GTGACACAGACACCGCGGTGATTT 901
196 uGlyPheGlySerLeuAlaGlyValGlnAlaValGlyGluIleTyr 213
902 CCGGTTCGATTTGAACGCG..GCGATTACACAGCGCGCGCATTAAT 948
213 rglIleuThrGlyAsnGlyValAlaIleIleAsnAlaGlnHisAsnHis 229
949 TTGGGACGTTACCAACATCAGATTCC.....GTTATCGAAGAAGCGCG 992
230 LeuAspIlePheIleGlySerIleSerAsnHisAlaIleGlnIleTyrPheAs 246
993 CACGAAAGAG.....C 1003
246 nGlyLysAspTyrSerAlaValAsnIleHisThrSerGlnIleTyr 263
1004 TGTGGCGTGGGTGCGCGCGCGCGCAAAATTAATCAGTACGCGTACA 1053
263 hrPheThrLeuIleSerProGln..GlyGluIleGluIleThrLeuPro 278
1054 ACCCTCGGC...CATTCCTGAAAACAACCTTCACAGTTCACACAGC 1100
279 TyrLeuGlyGlnHisAsnValLysAsnAlaLeuAlaThrAlaLeuAl 295
1101 CGTCAAC..GCGCGCAGCGCGCGCGCGCGCGATTGCTACTTACGAG. 1146
295 aMetAsnValGlyAlaThrLeuThrAspValLysAlaGlyLeuGln 312
1147 .....CGCGTATGCCCTTGATCCGATCCCGCCACCTG 1179
312 rGserGlnValLysGlyArgLeuPheProIleGlnValIleProAsnLeu 328
1180 CTTTTCGCGCATTTAATGTCGCGATACGACAGCGCGAGCGCATGGG 1229
329 LeuLeuLeuAsp.....AspThrTyrAsnAla..... 337
1230 TTGCTGGATTTGACGAGAACGCTTCGTTGTCAGCT 1270
338 .....AsnLysAsp..SerLeuGlyAlaAla 345
seq_name: swissprot_40:ATPB_PYLII
seq_documentation_block:
ID ATPB_PYLII STANDARD; PRT; 481 AA.
AC P65532;
DT 01-AUG-1992 (rel. 23, Created)
DT 01-AUG-1992 (rel. 23, Last sequence update)
DT 01-MAR-2002 (rel. 41, Last annotation update)
DE ATP synthase beta chain (EC 3.6.3.14).
GN ATPB.
OS Pytalsetta littoralis.

```

```

OG Chloroplast.
OC Eukaryota; stramenopiles; Phaeophyceae; Ectocarpales;
OC Acinetosporaceae; Pytalsetta.
OX NCBI_TaxId=2885;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92216062; PubMed=1532750;
RA Jouanne S., Kerbourg H.C., Kloareg B., Loiseaux-De Goer S.;
RT "Nucleotide sequences of the atpB and the atpE genes of the brown
  alga Pytalsetta littoralis (L.) Kjellm."
CC Plant Mol. Biol. 18:819-822(1992).
CC -! FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
  GRADIENT ACROSS THE MEMBRANE. THE BETA CHAIN IS THE CATALYTIC
  SUBUNIT.
CC -! CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate +
  H(+) (out).
CC -! SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
  CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
  SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
  HAS THREE MAIN SUBUNITS: A, B AND C.
CC -! SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
CC -! SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sdb.ch/announce/
  or send an email to license@isb-sdb.ch).
CC EMBL: X60329; CAA42899.1; -.
DR PIR: S17102; PMPFBL.
DR HSSP: P00829; IBMF.
DR Mendel: 2241; PYLI; atpB.1.
DR InterPro: IPR004100; ATP-synt_ab_N.
DR InterPro: IPR000793; ATPase_AB_C.
DR InterPro: IPR000194; ATPase_alpha_beta.
DR Pfam: PF00006; ATP-synt_ab; 1.
DR Pfam: PF00306; ATP-synt_ab_C; 1.
DR Pfam: PF02874; ATP-synt_ab_N; 1.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; 1.
DR ATP synthase; Chloroplast; Thylakoid; Membrane; CF(1);
  KW Hydroxylase; ATP-binding; Hydrogen ion transport.
  FT NP-BIND 161
  FT SEQUENCE 481 AA; 52025 MW; 9A270BAC2327416F CRC64;

```

```

alignment_scores:
  Quality: 104.50      Length: 513
  Ratio: 0.479        Gaps: 25
  Percent Similarity: 42.495  Percent Identity: 21.832

```

alignment\_block:

US-09-303-518D-125 x ATPB\_PYLII ..

Align seg 1/1 to: ATPB\_PYLII from: 1 to: 481

```

13 AAAAAGGTCTAAACCTGCCATCGCGGCGACGCGAGCAACCGCTTAA 62
  ::::::::::::::::::::
  9 GluYsglyIleAsn.....AsnGlyTyrIleThrGlnValIle.. 21
63 CGAGCGCGCGCATTTACCGAGTC.....GCGTTCGTTGGGGAAG 103
  |||||:::||||
22 ....GlyProValIleAspAlaValPheSerSerGlyIleLeuProLysI 37
104 AATATGCCGTATGCGCGCCCTCGATGAAGTCAAGAGGAGCGATGCCGCG 153
  |||:::||||
37 LetyrAsnAlaLeu.....GluValGlnSerLysLeuIleProIleIle 51
154 AAAAAGCGCAAGTGCCTTTGAGACGCAAAAGAAATCCGCGCGTGTGT 203
  ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
52 CysGluValGlnIleuLeuGlnLysAsnArgValArgAlaIleAlaIle 68

```

```

204 TACTGGCGGGCTTACGCAAAATCGCCGATTCACGTGGCGCAAAAGC 253
68 tSerAla.....ThraspIyleuGlnAlaGlyValThyV 80
254 GCGTACTGACGTCGTCGATTGCCGTTGAA..... 285
80 allleasprhInAlaProIleAlaValAlaValAlaValAlaThrLeu 96
285 ..... 285
97 G1YArG1LePheAsnValLeuGlyGlnProValAlaAspAsnLeuSerAspE 113
286 ....GGCAACGACGAAATCGATTGTTAGCGCTACGACCGCTAGACGCGTGG 331
113 rValGlyGlnAspPheLeuProIleHisArgSerAlaPro...AlaPheT 129
332 CAACCTTAAGCGCGGAGAAAGTGGCGCGCAACCTGATCCAAACCGGTTG 381
129 hrAspLeu.....GluThrLysProAlaLePheGluThrGlyLe 142
382 TGGACTGCGTGGCGACCCGCTCCGTCAGCAAAATTCCTGCGTGGATGC 431
143 LysValValAspLeuLeuAlaProTyrArgArg.....G1 154
432 CGAGCCGTCGCGCATCTTCGTCATCGCATGACACCAATCCGCTGGTG 481
154 yGlyLysIleGlyLeuPheGlyGlyAla.....GlyV 165
482 CCGACCGCTAGCGCATTCATCAAGAAAGCGCGCGAGATTCCAACGC... 528
165 a1GlyLysThrValLeuIleMetGluLeuIleAsnAlaLeAlaLysAla 181
529 .....GCGCTGTGTGATTAGCGCGTTGACGACGACGCAAAATCCATGT 572
182 HisGlyValSerValPheGlyGlyValGlyValGlyValGlyValGlyVal 194
573 TTGTAAAGGCACTGGCGGACGACGCGCGCTCGCAAAATGTCGCAACATCG 622
195 .....ThraArgGluGlyAsnAspLeuTyrMetG 204
623 AAACGACATGAATTCGGCGCGCGCGCATCGCGGTTTGAAGTGCACGAC 672
204 lUmetLysIleSerGlyValIleAsnGluThrAsnLeuLeuGluSerLys 220
673 ATTCATTTCATC.....GAGCGGTCGCGCGGATTAAC 707
221 ValAlaLeuValTyrGlyGluMetAsnGluProGlyAlaAlaMetAr 237
708 CGTGTG.....ACATCAATT 724
237 gValGlyLeuThrAlaLeuThrMetAlaGluTyrPheArgAspIleAsnL 254
725 ATCAAGATGATTAATAC...ATGGCGCTTGTGTTGCAACGGCGCGCTG 771
254 ySglnAspValLeuLeuPheIleAspAsnIlePheArgPheValGlnAla 270
772 AACACCGGACGCGGATTCGCTAGT.....GGTCTCAAGTCACAA 815
271 GlySerGluValSerAlaLeuLeuLysArgMetProSerAlaValAlaGlyT 287
816 ACCGCGCTTGTGCTACCGTTTGGTGGCGCAAGTATCCGAATTAATG 865
287 rGlnProThrLeuGlyThrGluMetGlyAlaLeuGlnGluThr 303
866 CGGGGGAATGTGTTGACACAGACACCGCGTATTCGTTGCGTATG 915
304 .....SerThrThr 306
916 AACGGCGCATTAACACAGCGCGGACGATTAATTTGGACGCTACCA 965
307 GlnGlySerIleThrSerIleGlnAlaValTyrVal..... 318

```

```

966 TCAGATTTCGCTTATCGAAGAGCGCGCAAAAGAGCTGTGGCGG 1015
319 .....ProAlaAspAspLeu.....T 324
1016 TTGGCGCGACCGCGACAAATACATCAACGCGCTACACGCGGCTG 1065
324 hrAspProAlaPro.....AlaThrThrPheAlaHis 324
1066 TTCCGAAAAAACAACGCTTCAGATTCAACAGACGCGCTCAACGGCGG 1115
335 Leu.....AspAlaThrThrValLeuSe 342
1116 CCGGCGCATGTGGCGCATGTGCTACTTACAGCGCGGTATGCGCTTG 1165
342 rArgGlyLeuAlaAlaLysGlyIleTyrProAlaValAlaAspLeuAsp 359
1166 TCGTCCCGACCGCTGCTTTCGCGCATTTAATCGCGCGAT..... 1206
359 erThrSerThr...MetLeuGlnProLeuIleValAlaGlyAspGluHisTyr 374
1207 .....AC 1208
375 LysThrAlaGlnLeuValLysGluThrLeuGlnArgTyrLysGluLeuG1 391
1209 GCACAGCGCGGACGCGATTCGCTTGGAAATGGACGACAAAGAC... 1254
391 nAspIleIleAlaIleLeuGlyIleAspGluLeuSerGluLysArgL 408
1255 .....CTGCGTTTGGCGACG 1269
408 euValValAspArgAlaArgLysIleGluArgPheLeuSerGlnProPhe 424
1270 TTCGTCGCG.....CCGGCGCAATAC 1290
425 pheValAlaGluValPheThrGlySerProGlyLysTyr 437
seq_name: SwissProt.40:PPSA_PYROHO
seq_documentation_block:
ID PPSA_PYROHO STANDARD; PRT; 821 AA.
AC 057830;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable phosphoenolpyruvate synthase (EC 2.7.9.2) (Pyruvate, water
DE dikinase) (PEP synthase).
DE PPSA OR PH0092.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=96344137; PubMed=9679194;
RA Kawarayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Toshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
RL DNA Res. 5:55-76(1998).
CC -I CATALYTIC ACTIVITY: ATP + pyruvate + H(2)O = AMP +
CC phosphoenolpyruvate + phosphate.
CC -I PATHWAY: ESSENTIAL STEP IN GLUCONEOGENESIS WHEN PYRUVATE AND
CC LACTATE ARE USED AS A CARBON SOURCE.
CC -I SIMILARITY: BELONGS TO THE PEP-UTILIZING ENZYMES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL, AF000001, BAA29161.1, -.  
DR HSSP, P22983, IDIK.  
DR InterPro, IPR000121, PEP\_utilizers.  
DR InterPro, IPR002192, PPK\_N\_term.  
DR Pfam, PF00391, PEP\_utilizers, 1.  
DR Pfam, PF02896, PEP\_utilizers\_C, 1.  
DR Pfam, PF01326, PPK\_N\_term, 1.  
DR ProDom, PDO00940, PEP\_utilizers, 1.  
DR PROSITE, PS00370, PEP\_ENZYMES\_PHOS\_SITE, 1.  
DR PROSITE, PS00742, PEP\_ENZYMES\_2, 1.  
KW transferase; kinase; ATP-binding; Phosphorylation; Complete proteome  
FT MOD\_RES 444  
FT DOMAIN 812 818  
SQ SEQUENCE 821 AA; 90812 MW; A93816D865F8A0BF CRC64;

|                     |             |
|---------------------|-------------|
| alignment_scores:   |             |
| Quality:            | 104.50      |
| Ratio:              | 0.437       |
| Percent Similarity: | 52.527      |
|                     |             |
|                     | Length: 455 |
|                     | Gaps: 21    |
| Percent Identity:   | 21.099      |

alignment\_block:

US-09-303-518D-125 x PPSA\_PYRHO .

Align seg 1/1 to: PPSA\_PYRHO from: 1 to: 821

```

107  GAGCCAGCCATATACCGAAGTCGCGCTTCCTGGCGAAGATATGCCGGTAT 116
125  GLyGluAlaValAlaSerGlyAlaValThrProAspLutryLLeuValG1 271
117  GCCGCCCTCCATGAAATCAAGGAGGCCGATGCCGCTAAAAAAGCCAAAG 166
271  uLySGlyThrTrpLysLLeuSGLu..LysAlIleAlaLysLysLuv 287
167  TGCCTGTTTGAAGCAAAAAGATCCGGCGCGTGTGTTTACGCGCGGCT 216
287  AlMet.....ValIleArgAsnProGlu 294
217  TCAGCGCAA.....ATGCGCGGATTCACCCGTGC..... 246
295  ThrGlyLysGlyThrValGlnValLysValAlaGluTrpLeuLysProG1 311
247  .....GAAAAGCGCGTACTT.....CACTGAGTCGATGCGCGGTG 283
311  uTrpValGluLysGlnValLeuThrAspLcuglnIleIleGluValAlaL 328
284  AAGCGACACAGCAAAATCGATTTGACGCGTACGCA..CGTGAAGCGCTG 330
328  ySmetLysLynLysIleGlu..GluHisTyGlyTrpProGlnAspIle 343
331  GCA..AACTTAAGCGCGCAAGAGAAGTGCGCCGCCAACCTGATTCATCCG 377
344  GluTrpAlaTyraSpsLysAspLysLysLeuTyrlIleValGlnSerAr 360
378  TTTGTGACGTGCGTCGCGCACCCGTCGCTGACGAAATTCCTGCGCGTCG 427
360  gProIleThrThrLeuLysGluThrThrThrGluGluValAlaGluGluVal. 376
428  ATGCGGACCGGTGCCCATCTTCCTGCATATGCGAGTGAACCAATCG... 474
377  ....GluGluAlaGluValIleLeuLysGlyLeuGlyAlaSerProGly 391
475  CTGCTGCCGCAACCTACGCGTATTATCAAAAGCGCGGAGATTCAAA 524
392  IlleGlyAlaGlyArgValValValIlePhesAspIleSerGlu..... 405
525  ACGGCGCTGTGGATATGAGACCGGTTGACCGCAAGCAAAATCCATGTTT 574
406  .....IleAspLysValLysGluGlyAspValLeuVal 417

```

```

575 GTAAGGACGCTGGGCGGAC...TGCGCGCT...GAAATTCGCGCAC 618
417 hTrhmetLtrhnsProAspMetValProIametyLsAlghalSeraIa 433
619 ATGGAACACATGAATTCGGCGGCCGACACGTCGCGGTTGAGTGAC 668
434 lIeIlelTrhAspIuGlYlArGrThSerhISaIaIaIleValSerA 450
669 GCACATTCATTCGATGAGCGCGTCGCGCGGATTAACACGCTGGACA 718
450 gLlueuGLYlIeProIaIeValIeGLYlThrLysGLuAlaThLysL 467
719 TCATATATCAAGATGAATTACCATTGGC..... 747
467 euLysThrGLYAspTYrValThIValAspLYlThrArGLYleuVal 483
748 .....CGTTGTTGGCAACGCGCGCTGAACACGAGCGCT 785
484 LysGLYlIeValLysSerIeValIGluLYlSLSYLSGLuIduIaI 500
786 GATTCCGCTAGTGCTTCAAGTCAACAACACGCGCTTGGCTGCAACG 835
500 aAlaIaProGLYIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 517
836 TTTTGGGTCCGAAGATATGCGCAATTACTCGCGCGCAATGCTTGACA 885
517 euValIeValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaI 529
886 GACAACCGCTGATTTCCGTTGCGTATTCGACGCGCGCATTAACAAG 935
530 .....ArgIaIaIaIaIaThrGLYlAspGLYlValIGLYleuIaI 544
936 CGCGACCATTAATTGGGACGCTACCAACAATCAGATTTCCGTTATGAG 985
544 aGLuIhISetIleuSerIleGLYlGluhISProValLysPheIleLys 561
986 AAGCGCGCAGAAGAAGCTGTCGCGTGGTGGCGCGCGCACGCGGACAA 1033
561 lUGLYlSGLUAspIuIeValIGluAlaGLYlIleGLYlS 577
1036 TACTCCATCACGCGTACCAACCTCGGCCATTTCTGAAAAACAACCTT 1085
578 ValIaIaIaIaIaPheTYrProArProVal...TrPTYrArGrThLeu 592
1086 CAATTCACAACACACGCTGCACAGCGGCGACGCGCGCATGTCGCCGATG 1135
593 AspIaIaProhIraNGluIuPheArGLIuIePro...GLYlSGLUAspI 608
1136 GTACTTCAGACGCGTATGCCCTTGATATCCGCGC..... 1172
608 uProGLUGluArGAsnProMetLeuIGLYlTPArGLYlIleArGrGLY 625
1173 .....CACCGTCGTTTGGCGGATTAATGCTGGCGCATGCGCACGCGC 1217
625 euAspGLIuProGLIuIeValIGluAlaIuPheLysAlaIleLysVal 641
1218 GCAAGCATTTGGTGGTCTGGAATTTGACGAAGAAGACCTGCTTTGCA 1267
642 ValIGluSGLYlUGAsnIleGLYl.....ValMe 652
1268 GCTTCGCTG.....CCGCGGCAATACGAATACGCGCGCGCTGTTG 1308
1309 CGCAAGATGCTGGA 1322
669 lAaRGLYlValIGly 673
seq_name: SwissProt_40:FIG2_YEAST
seq_documentation_block:
ID FIG2_YEAST STANDARD: PRT: 1609 AA.
```

P25653;  
 AC 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Factor induced gene 2.  
 GN FIG2 OR YCR089W OR YCR89W OR YCR1102.  
 OS *Saccharomyces cerevisiae* (Baker's yeast).  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OC NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92397594; PubMed=1523889;  
 RA Wilson C., Grisanti P., Frontali L.;  
 RT "The complete sequence of a 6146 bp fragment of *Saccharomyces*  
 RT *cerevisiae* chromosome III contains two new open reading frames."  
 RL Yeast 8:569-575(1992).  
 CC - FUNCTION: REQUIRED FOR EFFICIENT MATING.  
 CC - INDUCTION: BY MATING PHEROMONES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X59720; CAA42254.1; -  
 DR PIR: S19504; S19504.  
 DR PIR: S25345; S25345.  
 DR SGD: S0000685; FIG2.  
 SQ SEQUENCE 1609 AA; 166049 MW; 7D66AD/F85A/B852 CRC64;

alignment\_scores:  
 Quality: 104.50 Length: 476  
 Ratio: 0.475 Gaps: 19  
 Percent Similarity: 46.218 Percent Identity: 19.958

alignment\_block:  
 US-09-303-518D-125/rev x FIG2\_YEAST ..

Align seg 1/1 to: FIG2\_YEAST from: 1 to: 1609

```

1331 TCAAGTTTCCAGCAGCTTTCGCAACAGCGCGCGCTGATTCGATTGGC 1282
      ::::::::::::::::::::: |||::: |||::: |||:::
982 ThrValIleThrSerValGlnMetArgSerThrProPheProIleuThr 998
1281 CGGCGACAGCAAGCTGCACAAAGCGAGGCTTCTGTCCTCAATTCACAGC 1232
      ::::::::::: :: |||::: ::::::::::: |||:::
988 rSerSerThr.....SerSerSerSerLeuAlaSerThrIleu 1011
1231 AACCCAAATGCTGGCGGCTGTGGATTCGCGGACGAGATTAAATGCGCGCAA 1182
      ::::: ||| ||| |||:::
1011 ySerSerLeuGlnAlaSerSerGlnMetSerThrPhe..... 1023
1181 ACCAGGTTGGGAGAGATTCACAGGCGATCAAGCGCTCGTAAGTACCAAT 1132
      ::::::::::: ::::::::::: |||:::
1024 .....SerValSerThrGlnSerLeuProle 1032
1131 CGGC...ACCATGGCGCGGTGCGCGCTTACAGCGCTGTGTGAACCTGA 1085
      ::::: ||| ::::::::::: |||::: |||:::
1032 uAlaPheThrCysSerGlnLysArgSerThrThrSerVal..... 1045
1084 AGAGTTTGTTCAGAAATGCGGAGGTTTACAGCGGTGATGAGTAT 1035
      ::::: ||| ::::::::::: |||:::
1046 .....SerGlnTrpSerAsnThrValLeuThrAsnThrIle 1057
1034 TTGTCCGGTGGCGGCAACCCAGCGCAAGCTCTTGTGCGGCGCTTC 985
      ::::: ||| ::::::::::: |||:::
1058 MetSer.....SerSerSerAsnValIleSerThrAsnGlnLysProSe 1072

```

```

984 TTTCGATACCGGAATCTGATGTGTAGCTCCCAAAATTAATCGCGCGC 935
      ||| |||
1072 rSerThrThr..... 1075
934 CTTGTGTATTCGCGCGTTCAT..... 912
1076 .....SerProIleAsnPheSerSerGlyTyrSerLeuProSer 1088
911 ...ACCGAACCGGAATCAACGCGGTTCGTGTGTCACAC.....AA 874
      ::::: |||::: |||::: |||::: |||::: |||:::
1089 SerSerThrProSerGlnTyrSerLeuSerThrAlaThrThrIleAs 1105
873 TTGCGCCGCGATTAATTTGGATCTTTCGCCAACCAACGCGTACGACAGA 824
      ::::: |||::: |||::: |||::: |||::: |||:::
1105 nGlyIleLysThrValTyrThrThrTyrCysProLeuAlaLysSer 1121
823 GCGCGCGTTTGTGACTTTCAGACCACTAGGCGCATACGCGCGCGGTG 774
      ||| |||
1122 .....ThrValAlaAlaSerSerGlnSerSerArgSerVal 1133
773 TTTCAGCGCGCTGTTCACAAACAGCGCAATGTAAATTCATCTTGATA 724
      ||| |||
1134 AspArgPheValSerSerSerLys..... 1141
723 ATTGATGTTCACACGCTTTTATTCGCGCGCGCGCTCGAATGAATGA 674
1142 .....ProSerSerLeu..... 1146
673 TGTGCGTGCACCTCAACACCGCAGATCGGCGCCCGCAATTCATGTGT 624
      ||| |||
1147 .....SerGlnThr 1149
623 TCGATGTTCGACGATTTTCAGACGCGACGCTGCGCGCGCTTCACA 574
      |||::: |||::: |||::: |||::: |||::: |||:::
1150 SerIleGlnTyrThrLeuSerThrAlaThrThrIleSerGlyLeu 1166
573 A.....ACATGATTTTTCGCTTCGCGTCAACAGCGCTCATATACA 536
      ||| |||
1166 sThrValTyrThrThrTyrCysProLeuThrSerLysSerThrLeuGlyA 1183
535 ACAGCGCGCGTTTGAAATCTCGCGCGGCTTCTTGATTAATACCGTAGGG 486
      ::::: |||::: |||::: |||::: |||:::
1183 LaThrThrGlnThrSerSerThrThrAlaLysValArgIleThrSerAlaSer 1199
485 TCGCACCGACGCGATTTGCTGCTCATCCATTCACAAAGTGGCAACGG 436
      |||::: |||::: |||::: |||::: |||:::
1200 SerAlaThrSerThrSerIleSerLeuSer.....ThrSerThrG1 1213
435 CTCGCATTCGACGACGAGAAATTTTCGTCGACGAGCGGTCGCGACGCGCAG 386
      ||| |||::: |||::: |||::: |||:::
1213 uSerGlnSerSerSerGlyTyrLeuSerLysGlyValCysSerGlyThrG 1230
385 TCACAAACCGGATTTGATGATCAGTTGCGGCGCACTTCGCGCGCTTAG 336
      ||| |||
1230 LucyThrGlnAsp.....ValProThrGlnSerSerProAlaSer 1244
335 TTTCGACGCGCTTCAGGTCGCTAGCTTCACAACTCGATTCGTCGTTGCC 286
      ::::: |||::: |||::: |||::: |||:::
1245 ThrLeuAlaTyrSerProSerValSerThrSerSerSerSerPheSe 1261
285 TTCAACGGCAATCAACACTGACTGAACTACGCGCTTTCGCCACGCTGAA 236
      ::::: |||::: |||::: |||::: |||:::
1261 rThrThrThrAlaSerThrLeuThrSerThrHisThrSer..... 1274
235 TCGCGCGCATTTTTCGCGAACGCGCGCAGTAAACAC..... 198
      ::::: |||::: |||::: |||::: |||:::
1275 ..ValProLeuProSerSerSerSerSerIleSerAlaSerSerProser 1290
197 .....ACGCGCGAATCTTTTGTGTC 178
1291 SerThrSerLeuLeuSerThrSerLeuProSerProAlaPhe..... 1304
177 TTCAAAACAGCACTTGCGCTTTTTCAGCGGCAATGCCCTTCTTACATTCA 128

```



```

1305 .ThrsSerSerThrlProThrlAlaValSerSerThrPheI 1321
127 TCGAGGGGGCATACCGGATATTCT.....TCGCGA 96
1321 LeAlaSerSerLeuProLeuSerSerSerSerSerLeuSerPro 1337
95 ACCAACGGCAGCTTCGTAATG.....GCCGGCGCGTGTGA 61
1338 ValSerSerSerIleuMetSerIlePheSerSerSerSerSer 1354
60 AACGGCTTGCTCCGCTTCGCCCGCATG 33
1354 rSerSerLeuAlaSerLeuProSerLeu 1363

```

seq\_name: SwissProt\_40:ATPB\_DICDH

seq\_documentation\_block:

ID ATPB\_DICDH STANDARD; PRT; 481 AA.

AC P30158;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE ATP synthase beta chain (EC 3.6.3.14).

GN ATPB.

OS Dictyofa dichotoma.

OG Chloroplast.

OC Eukaryota; stramenopiles; Phaeophyceae; Dictyotales; Dictyotaceae;

OC Dictyofa.

NCBI\_TaxID=2876;

RP SEQUENCE FROM N.A.

RA MEDLINE=92322956; PubMed=1535802;

RT Leitsch C.E.W., Kowalik K.V.;

RT "Nucleotide sequence and phylogenetic implication of the ATPase

subunits beta and epsilon encoded in the chloroplast genome of the

Plant Mol. Biol. 19:289-298(1992).

RT FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON

GRADIENT ACROSS THE MEMBRANE. THE BETA CHAIN IS THE CATALYTIC

SUBUNIT.

CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(in) - ADP + phosphate +

H(+)(out).

CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC

CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE

SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)

HAS THREE MAIN SUBUNITS: A, B AND C.

CC -1- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.

CC -1- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: X66939; CAA47370.1; -

DR PIR: S22509; S22509.

DR HSSP: P00829; IBMF.

DR Mendel: 2225; DICDH:atpb.1.

DR InterPro: IPR004100; ATP-synt\_ab.N.

DR InterPro: IPR000793; ATPase\_AB.C.

DR InterPro: IPR00194; ATPase\_alpha\_beta.

DR Pfam: PF00006; ATP-synt\_ab.1.

DR Pfam: PF00306; ATP-synt\_ab.C.1.

DR Pfam: PF02874; ATP-synt\_ab.N.1.

DR PROSITE: PS00152; ATPASE\_ALPHA\_BETA.1.

DR ATP synthase; Chloroplast; Thylakoid; Membrane; CF(1);

Hydroxylase; Hydroxylase; Hydrogen ion transport.

NP\_BIND 161 168 ATP (BY SIMILARITY).

SEQUENCE 481 AA: 52203 MW; 849B15C113FD1F91 CRC64;

# alignment\_scores:

Quality: 103.50 Length: 515  
Ratio: 0.502 Gaps: 25  
Percent Similarity: 40.000 Percent Identity: 21.942

## alignment\_block:

us-09-303-518d-125 x ATPB\_DICDH ..

Align seg 1/1 to: ATPB\_DICDH from: 1 to: 481

```

67 GCGCGGCGCATATACCGAGTCGCTTCGTCGCGAAGATATGCGGATAT 116
|||||:|||||
21 GlyProValIleAspAlaValPhe.....SerIleGlyI 32
117 GCGGCGC.....TCGATGAAGTCAGAGAGCGGATGCC.... 150
: ||| :|||:|||||:|||||:|||||:|||||:|||||:
32 nleuProLysIleTyrAsnAlaLeuGluValLysSerLysAspGlyThr 49
151 .....GTCAAAAGAGCGCAAGTGTGTTGAACAGCAAAAGATCCGGC 195
||| :|||:|||||:|||||:|||||:|||||:|||||:
49 hrlleIleCysGluValGlnGlnleuPheAsnAspAsnArgValArgAla 65
196 GTGCTGTTACTCGCGCGCGCTTCAGCAAAATCGCGCGATTCACCGTGG 245
:|||||:|||||:|||||:|||||:|||||:|||||:
66 IleAlaMetSerIle.....ThrAspGlyLeuGlnArgI 77
246 CGAAGAG..... 252
| :|||:|||||:|||||:|||||:|||||:
77 YValGluValIleAspThrGlnAlaProIleLeuValProValGlyLysA 94
253 .....CGCGTACTTCAGTCACTGTCGATGTCGCGCT 282
:|||||:|||||:|||||:|||||:|||||:|||||:
94 latThrleuGluArgIlePheAsnValLeuGlyGlnThrValAspAsnIle 110
283 GAA.....GCGAAGAGCAAGATTCAGATTGAACGCTACGACATGA 323
||| :|||:|||||:|||||:|||||:|||||:|||||:
111 GlnIleGlyThrGlyLysArgIleProIleAsnArgProAlaProse 127
324 AGCGCTGCGCAACTGAAGCGCGAAGAGTGGCGCGCAACATGCATCAAT 373
: |||:|||||:|||||:|||||:|||||:|||||:
127 r..... 127
374 CGCGTTTGAGACTGCGCTGCCACCGCTCCGTCGAGCAAAATTCCTGCC 423
:|||||:|||||:|||||:|||||:|||||:|||||:
128 .....PheThrAspLeuGluThrLysPro..... 135
424 GTTCGATGCCAGCGCTTCGCGCATTCGTCGAATGGATGAGACCAATCC 473
|||:|||||:|||||:|||||:|||||:|||||:
136 .....AlaIlePheGluThrGlyIleValValAs 146
474 GCTGCTGCC..... 483
|||:|||||:|||||:|||||:|||||:|||||:
146 pleuLeuAlaProTyrArgArgGlyLysIleGlyLeuPheGlyLysA 163
484 .....GACCTACGTCATATACCAAGAAAGCGCGAGGATTTGAA 525
|||||:|||||:|||||:|||||:|||||:|||||:
163 IagIValGlyLysThrValLeuIleMetGluLeuIleAsnAlaIleAla 179
526 GCG.....GCGCTGTGTATGAGCGGTCGTCGAGCAAGCAAAAT 566
:|||||:|||||:|||||:|||||:|||||:|||||:
180 LysAlaIleGlyLysValSerValIlePheGlyLysValGlyLysArg.... 194
567 CCATGTTTGTAAAGCGAGCTGCGCGAGAGCGCTGTCGAAGATGCTGCCA 616
:|||||:|||||:|||||:|||||:|||||:|||||:
195 .....ThrArgGluGlyAsnAspLeu 202
617 ACATGAAACACATGAATTCGCGCGCGCGCGATTCGCGGTTGATGCGG 666
:|||||:|||||:|||||:|||||:|||||:|||||:
202 ymeGluMetLysGluSerLysValIleAsnGluSerAsnLysSerGlu 218
667 ACGCATTCATTCATTC.....GAGCGGCTGCGCGCGAA 701

```

```

::: ::: ::::: ||||| |||||:
219 SerLysValAlaLeuValTyrGlyGlnMetAsnGlnProGlyAlaLar 235
702 TAAACCGTGG.....ACCA 718
235 gmetatgValGlyLeuThrAlaLeuThrMetAlaGluTyrPheAlaGAspI 252
719 TCATTTATCAAGATGTAATTACC...ATTGGCCGTTTGGTTTGCACAGCC 765
252 IeAsnArgGlnAspValLeuLeuPheIleAspAsnIlePheArgPheVal 268
766 CGTGTGACACCGGCGCGTGGTGGCTAGT.....GGTTCCTCAAGT 809
269 GlnAlaGlySerGlyValSerAlaLeuLeuGlyArgMetProSerAlaVa 285
810 CACAAACCGCGGCTTGGCTGACCGTTTGGTGGGAAAGATACGCAAA 859
285 IGlTyrGlnProThrLeuGlyThrGlnMetGlyAlaLeuGlnGluArgI 302
860 TTAAGTGGGCGAATTGGTTGACACAGACACCGCGTGGTGGTGG 909
302 IeThr.....Ser 304
910 GATATGACGCGCGGATTTACACAGCGCGACGATTATTTGGACGCTA 959
305 ThrThrGlnGlySerIleThrSerIleGlnAlaValTyrVal..... 318
960 CCACAATCAGATTCGTTATCGAAGAGCGCGACGCAAGACGCTGGC 1009
319 .....ProAlaAspAspLeu..... 323
1010 GGTGGTGGCGCGCGCGGCAAAATATCTCATCAGCGCTACACCGCTC 1059
324 .....ThraSproAlaPro.....AlaThrThrPhe 332
1060 GGCCATTCCTGAAAAACAACTCTCAAGTTCAACACAGCGCTCAAGG 1109
333 AlaHisLeu.....AspAlaThrThrVa 340
1110 CGGCGACGCGCGCATGTCGCGATGCTAGCTAGACGCGTGTGGCTG 1159
340 ILeuSerThrArgGlyLeuAlaAlaLysGlyIleTyrProAlaValAspProL 357
1160 TGGATTCCTGCCACCTGCTTGGCGGATTTAATGTCGGCGGATACC 1209
357 euAspSerThrSerThr...MetLeuGlnProValIleValIleGlySerGlu 372
1210 .....GACAGCGCGCAG..... 1221
373 HisTyrAspThrAlaGlnLeuValLysLysThrLeuGlnArgTyrLysG1 389
1222 .....GCATTGGGTTGGCTGGATTGGACGAGAG 1252
389 uLeuGlnAspIleIleAlaIleLeuGlyIleAspGlnLeuSerGlnGlu 406
1253 AC.....CTGGCTTTG 1263
406 sprArgLeuValAlaAspArgAlaArgLysIleGluArgPheLeuSerGln 422
1264 TGCAGCTTGCTGTC.....CGGGCAAAATAC 1290
423 ProPhePheValAlaGluValPheThrGlySerProGlyLysTyr 437
seq_name: SwissProt_40:VL2_HPV31
seq_documentation_block:
ID VL2_HPV31 STANDARD: PRT: 466 AA.
AC P17389;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Minor capsid protein L2.
GN L2.

```

```

OS Human papillomavirus type 31.
OC Viruses; dsDNA viruses; no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10585;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89299478; PubMed=2545036;
RA Goldsborough M.D., Dissvestre D., Temple G.F., Lorincz A.T.;
RT "Nucleotide sequence of human papillomavirus type 31: a cervical
RT neoplasia-associated virus."
RL Virology 171:306-311(1989).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC -----
CC EMBL: J04353; AAA6955.1; -.
CC PIR: H3244; P2WL31.
CC InterPro: IPR000784; late_L2.
CC Pfam: PF00513; late_protein_L2; 1.
CC Coat protein, late protein.
KW SEQUENCE 466 AA; 49975 MW; 99AD125BC65B1368 CRC64;
SQ

```

```

alignment_scores:
Quality: 102.50 Length: 368
Ratio: 0.614 Gaps: 21
Percent Similarity: 45.380 Percent Identity: 23.641

```

```

alignment_block:
US-09-303-518D-125 x VL2_HPV31 ..

```

```

Align seg 1/1 to: VL2_HPV31 from: 1 to: 466

```

```

354 GCGCGCACCTGATCCAAATCCGTTGTGGACTGGCGCTGCGACCGCTC 403
119 AlaProAlaProIlePro.....HisPro.P 127
404 CG.....TTCAGCAAAATTCCTGCGCTGATGCGGACCGCTTC 441
127 roThrThrSerGlyPheAspIleAlaThrThrAlaAspThrThrProAla 143
442 GCCATCTTGTCAATGCGATGACAC.....AATCGCTGGCTGGCGA 485
144 ILeuAspValThrSerValSerThrHisGluAsnProThrPheThrAs 160
486 CCTTACGCTCATTTACAAAGAGCGCGGAGATTTCAAAGCGGCGCTGT 535
160 ProSerValLeuGlnProProThrProAlaGluThrSerGlyHisLeuL 177
536 TGGTATTCAGCGGTTTGACCGACGCAAAATCCATGTTTGTAAAGCAGCT 585
177 euLeuSerSer.....SerSerIleSerThrHisAsnGlnLysGluIle 191
586 GCGCGCAGAC.....GTCCGCTGAAATGCTCCCAATC.....GA 623
192 PrometAspThrPheIleValSerThrAsnGlnLysIleThrSerSe 208
624 AACACATGAATTCGCGCGCGCATCTCGCGGTTTG..... 660
208 rThrProIleProGlyValArgArgProAlaArgLeuGlyLeuTyrSerL 225
661 ..AGTGGCAGCATTCATTCATCGACCG..... 690
225 yAlaThrThrGlnGlnValLysValIleAspProThrPheLeuSerAlaPro 241
691 .....GTGCGCGGAA 701
242 LysGlnLeuIleThrTyrGluAsnProAlaTyrGluThrValAsnAlaG1 258

```

```

702 TAAACCGTGTGGACCATCAAT..... 723
258 ucluseuerytyrheserasthserhisasnllealaproasproa 275
724 ..TATCAAGATGTAATACATGGCCGTTGTTGCAACGAGCGGCTG 771
275 sspheleuaspillellealeuhsatgproalaleuthseratg 291
772 AACACCGAGCGCGTATGTCCTAGCGTGTCTCAAGTCAACAACCG 821
292 asnthrvalatgtyrseratglauglyasnysglnthrleuathr 308
822 CCTCTTGGGATACCGTTTGGGCAAGTA.....T 853
308 gserglyalathrile...glyalathrvalhsttyttrtyasples 324
854 CGCAATATCT...GCGGCGCAATGTTGTACACAGCAACCGCGTAT 900
324 erseerleasnproalagllyserleleuethleuethleuagly 340
901 TCCGTTGCGTATGTAACGCGCATACACAGCGCGCATATATT 950
341 Serlatrthr...thserthrleuasnspglyleuatyasp.ile 356
951 GGGAGCGTACC...ACAATCAGATTTCG.....TATCG 982
356 yralaasprthrspthethrvalasprthrproalathrhisasnl 372
983 AAGAGCGCGCGAGCAAGACTGTGCGGCGGTTGCGCGCGCGAGC 1032
373 Proserthrvalaiglnserthrseralavalseralatyvalpro 389
1033 AATATCT..... 1039
389 rasnthrthrvalproleuserthrglypheasplleprolephes 406
1040 .....CATCAGCGCTACACCGCTCGGCC 1063
406 lyproasprvalproleuhsalaprothrvalpheprothro 422
1064 ATTTCCTGAAAAACAACCTTCAGTTCACACAGCGCGTCAAGCG 1113
423 leu.....Alaprothrthrproglinalser 432
1114 GACCGGCGCATGTCGCGATGCTACTAGAGCGCGTATCGCA 1163
432 epheval.Aspglylyasprpethrtyl..... 441
1164 TATCTGCCCGACCTGTTTGGCGATTTATCGTGGCGCATACCG 1211
442 .....Hisproserthytyrmetleuysatgargatgylasg 454
seq_name: SwissProt_40:RNFC_PASMU
seq_documentation_block:
ID RNFC_PASMU STANDARD; PRT; 835 AA.
AC Q9CNP2;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Electron transport complex protein rnfC.
GN RNFC OR FMO385.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxId=747;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
"Complete genomic sequence of Pasteurella multocida Pm70.";

```

```

RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -1- FUNCTION: May be part of a membrane complex involved in electron
CC transport (By similarity).
CC -1- COFACTOR: Binds 2 4Fe-4S clusters (potential).
CC -1- SUBUNIT: Composed of at least six subunits; rnfA, rnfB, rnfC,
CC rnfD, rnfE and rnfG (By similarity).
CC -1- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -1- SIMILARITY: BELONGS TO THE 4Fe-4S BACTERIAL-TYPE FERREDOXIN FAMILY.
CC RNFC SUBFAMILY.
-----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AE006074; AK02469.1; -
DR InterPro; IPR001450; 4Fe4S_ferredoxin.
DR InterPro; IPR001949; Complex1_51k.
DR Pfam; PF01512; Complex1_51k; 1.
DR PROSITE; PS00198; 4Fe4S_FERREDOXIN; 2.
DR PROSITE; PS00198; 4Fe4S_FERREDOXIN; 2.
KW Electron transport; Iron-sulfur; 4Fe-4S; Inner membrane;
KW Complete proteome.
FT METAL 377 377 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 380 380 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 383 383 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 387 387 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 416 416 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 419 419 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 422 422 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 426 426 IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ SEQUENCE 835 AA; 90851 MW; E682CAAF7A19E79 CRC64;

alignment_scores:
Quality: 102.50 Length: 447
Ratio: 0.477 Gaps: 18
Percent Similarity: 48.098 Percent Identity: 19.463

alignment_block:
us-09-303-518d-125 x RNFC_PASMU
Align seg 1/1 to: RNFC_PASMU from: 1 to: 835
103 GAATATGCCGCGTATGCGCCCTCATGAAGTCAAGGCGATGCCGT 152
48 GlnHisAlaGlyLysAlaGlyValHisValGlyAspTyrVal 64
153 CAATAAGCGCCCAAGTCGCTGTTGAAGACAAAGAACGCGCGTGT 202
64 PheLysGlyGlnProLeuThrGlnGlyAspGlyLeuArgValLeuPro 81
203 TTAATCGCGCGCGCTTCAGCAAAATCGCGCGAT.....CACCGT 243
81 AlnHisAlaSerThrSerGlyPheLeuAlaGlyAlaLeuProTyrAlaSer 97
244 GCGGGAAGAGCGCGTCTTCAGTCAAGTCGATTCGCGTGAAGCAACA 293
98 AlaHisProSerGlyLeuAlaThrLeuLysLeuHisValGlnLysAla 114
294 CGAATCGAGTTTGAACGCTACGACCT...GAAGCGCTGCAAACTTAA 340
114 yLysAspGlnThrArgGlnGlnGlnProLeuAspAspThrLeuThrGln 131
341 GCGGGAAGAGAGTCCGCGCAACCTGATTCACATCGGTTGTGACTGCG 390
131 hrProGlyLysLeuIleGlyLysLeuTyrGlnAlaGlyVal...AlaGly 146
391 CTGCGCAACCGCTCGCTTC.....ACGAAATTCCTGCGCTGATGC 431

```

```

147 LeuGIyGIyAlaValPheProThrAlaAlaLysLeuHisSerAlaGluLy 163
432 CGAGCGCTCGGCATCTTCGATCGATGGACACATCCGCTGGCTG 481
163 sgnValLysLeuLeuIleIleAlaGluLysGluProTyrIleT 180
482 CGGACCCCTAGCTATTCAGAAAGCCGCCGAGATTTCAGACGGCG 531
180 hrcysAspAspArgLeuMetArgSerTyrAlaAspGluIleLeuGlu 196
532 CTGTGGTGTAGTGGCGCTTGGACCGCAAAATCCATGTTGTAAGC 581
197 ThrArgIleLeu...ArgTyrIleLeuArgProGluLysValAlaIle 212
582 AGCTGGCGGACAGCTGCGCT... 604
212 aValGIuAspAsnLysProGluValAlaAlaLeuArgGluValLeuG 229
605 AAAATGCTGCCACATCGAAACATGATTCGGCGCGCGCATGCTGCC 654
229 lngIyAlaAsnAspIleGluIleArgValIleProThrLysTyrPro 245
655 GGTGGTGGACGCGCATTCATTCATGAG...CCGCT 692
246 GIyAlaAlaLysGluLeuIleGluIleLeuThrGlyMetGluValPro 262
693 CGGCGCGCATTAACCCGTGGACCATTCATTCAGATTCATTCAC 741
262 rGIyLArgSerSerSerIleGlyValLeuMetGlnAsnValAlaIle 279
742 ...ATTGGCGCTTGTGGCAACAGCGCTCGCAACAGCGCG 783
279 lApheAlaValLysArgAlaIleMetAspAspLeuProLeuIleGlu 295
784 GTGATTCGCTAGCTGCTTCATCAAGTCAACAAACCGCGCTTCGCT 833
296 ValValThrLeuThrGlyAspLysValArgHisLysGluAsnTyrTrp 312
834 CGTTGGTGGCGAAAGTATCGCAATTCATTCGGCGCGCATTCGTTG 883
312 lArgLeuLysThrProIleTyrGlnLeuLeuGln... 323
884 CAGACACCGCGGTATTCGCGTATTCAGCGCGCATTCACAA 933
323 ... 323
934 GCGCGCGCATTCATTCGGACGCTACACATTCATTCGTTACGA 983
324 ...GlnValAspTyr...HisTyrAspAspArgPheProValPhe 337
984 AGAAGCGCGCGCAAGAGCTGTCGCTGGCTGGCGCGCGCGCGGACA 1033
337 tGIyLeu...PrometMetGlyPheIleLeuProAspLeu...G 350
1034 AATATCTCCATCGCGGTACACCTCGGCGCATTCCTGAAAAACAATC 1083
350 lAlaIleProValThrLysMetThr...AspCysLeu 360
1084 TTC...AAGTCAACACGCGCTCAACGCGCGCGCGCGCAT 1124
361 LeuAlaProAspHisPheGluTyrAlaProProAlaProGluInSer 377
1125 GGTGCGCATTCGCTTACAGCGCGGTATGCGCTTGGATATCTGCCCA 1174
377 sIleArgCysSerAlaCysSerAspAlaCysProValSerLeuMetPro 394
1175 CCCTGCTTTGGCGCATTAATCGTCGGGATACGAGCGCGGAGCA 1224
394 lngInLeuTyrTrpPheAlaArgSerGluAspHisGluLysSerIle 410
1225 TTGGGT...TGCTGGAATTCAGCAAGAGCAAGCACTTCGTTG 1265
411 TyrAlaLeuLysAspCysIleGlu...CysGlyLeu 422

```

```

1266 CAGCTTCGCTCGCGCGCAATACGATACGCGCGCTGTGGC... 1311
422 sAlaTyrAlaCysPro...SerHisIleProLeuIleGluTyrP 436
1312 ...AAGTCTCGCAACCATTCGAGAG 1335
436 hArgGlnGluLysAlaLysIleTrpGluIleGlnGluLys 449
seq_name: swissprot_40:N121_HUMAN
ID N121_HUMAN STANDARD; PRT; 1229 AA.
AC Q9Y2N3; 075115; Q9Y4S7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nuclear envelope pore membrane protein POM 121 (Pore membrane protein
DE of 121 kDa) (P145).
GN NUP121 OR KIA0618.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
PI [1]
RP SEQUENCE FROM N.A.
RA Cordes M., Bauer C., Holmes A.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE OF 243-1229 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98403680; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
RL DNA Res. 5:165-176(1998).
RN [3]
RP SEQUENCE OF 1130-1229 FROM N.A.
RC TISSUE=uterus;
RA Koehler K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ESSENTIAL COMPONENT OF THE NUCLEAR PORE COMPLEX. THE
CC REPEAT-CONTAINING DOMAIN MAY BE INVOLVED IN ANCHORING COMPONENTS
CC OF THE PORE COMPLEX TO THE PORE MEMBRANE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. NUCLEAR PORE
CC MEMBRANE (BY SIMILARITY).
CC -1- DOMAIN: CONTAINS X-F-X-F-G REPEATS.
CC -1- SIMILARITY: THE REPEAT REGION COMPOSED OF PENTAPEPTIDE REPEATS
CC SEPARATED BY SER/THR-RICH DOMAINS IS SIMILAR TO THAT OF YEAST
CC NUP1, NUP1 AND MAMMALIAN P62 AND NUP153.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch)
CC -----
DR EMBL; AC006014; AAD28064.1; -
DR EMBL; AB014518; BAA31593.1; -
DR EMBL; AL080109; CAB45713.1; -
KM Nuclear protein; Transport; Transmembrane; Repeat.
FT DOMAIN 1 40 CISTERNA SIDE (POTENTIAL).
FT TRANSMEM 41 61 POTENTIAL.
FT DOMAIN 62 1229 PORE SIDE (POTENTIAL).
FT DOMAIN 4 10 POLY-ALA.
FT DOMAIN 51 56 POLY-ALA.
FT DOMAIN 294 299 POLY-LYS.
FT DOMAIN 441 444 POLY-SER.
FT DOMAIN 499 502 POLY-PRO.

```

```

FT DOMAIN 733 736 POLY-SER.
FT DOMAIN 819 826 POLY-SER.
FT DOMAIN 869 875 POLY-SER.
FT DOMAIN 1061 1067 POLY-SER.
FT CONFLICT 297 297 K -> E (IN REF. 2).
FT CONFLICT 336 336 S -> N (IN REF. 2).
FT CONFLICT 379 379 T -> A (IN REF. 2).
FT CONFLICT 451 451 K -> R (IN REF. 2).
FT CONFLICT 545 545 T -> I (IN REF. 2).
FT CONFLICT 559 559 S -> P (IN REF. 2).
FT CONFLICT 702 702 T -> P (IN REF. 2).
FT CONFLICT 881 881 P -> H (IN REF. 2).
FT CONFLICT 965 965 T -> A (IN REF. 2).
FT CONFLICT 982 982 PAT -> A (IN REF. 2).
FT CONFLICT 991 996 ASTRI -> PSMIRV (IN REF. 2).
FT CONFLICT 1000 1000 H -> Y (IN REF. 2).
FT CONFLICT 1006 1008 OPT -> HPI (IN REF. 2).
FT CONFLICT 1044 1044 MISSING (IN REF. 2).
FT CONFLICT 1102 1102 A -> T (IN REF. 2).
FT CONFLICT 1124 1124 T -> A (IN REF. 2).
FT CONFLICT 1146 1146 S -> G (IN REF. 1).
FT CONFLICT 1165 1165 L -> Q (IN REF. 1).
FT CONFLICT 1195 1195 A -> G (IN REF. 1).
FT CONFLICT 1199 1229 SHAPSPSIGSGKTPGARORLOARORHRRK -> NTFPHO
(OHSPRKGPNNLSKRKLIPAVRAAGPPRRGASSTPRKE
(IN REF. 2).
P -> L (IN REF. 3).
SQ SEQUENCE 1229 AA: 125087 MW: 515655D1285898B CRC64;

```

## alignment\_scores:

```

Quality: 102.50 Length: 457
Ratio: 0.539 Gaps: 18
Percent Similarity: 41.575 Percent Identity: 21.225

```

## alignment\_block:

```

US-09-303-518d-125/rev x N121_HUMAN

```

```

Align seg 1/1 to: N121_HUMAN from: 1 to: 1229

```

```

1313 TTGCGCAACGCGCGCGGCGTATTCGTAATTTGCCGCGGCAAGCAAGCTGCA 1264
      ||| ::||| ||| ::|||
616 LeuSerGlnSerGlyProGlyLeuLeuPro..... 626
1263 CAAGAGGAGGTCTTCTGCTCCATTCACAGCAACCAATGCC...TCGC 1217
      ||| ||| ::||| ||| ::|||
627 .....SerProSerPheAspSerIleProThrThrLeuLeuG 640
1216 CGCTGCGGTATCGCGCAGATTAATCGCGCAAAAGCAGGTTGGCGCAGG 1167
      ::||| ::||| ::||| ::|||
640 ILeuIleProAlaProSerMet..... 647
1166 ATATCCAAAGGCAATCAAGCGGCTGTAAGTACCAATCGCACCAATGCGCG 1117
      ||||| ::||| ::||| ::|||
648 .....ValPro...AlaThrAspThrIle 654
1116 GTGCGCGCGGTGAGCGGTGTTGAATTAAGTATGTTGTTTCAGGA 1067
      ::||| ::||| ::||| ::|||
654 AlaLeuProThrIleGlnAlaGluThrAlaThrIleProGlnAlaThrS 671
1066 AATGCGCGAGGTTGTACGCGTGAATGATTTGTGCGGCTGGCGGCA 1017
      ::||| ::||| ::||| ::|||
671 eAlaIleProSerProAlaProIleGlnSerPheLeuPheGlyThrGlnAsn 687
1016 ACCGACGCGAAGAGCTCTTGTGCGGCTCTTCGATTAACGAAATCTG 967
      ||| ::||| ::||| ::|||
688 ThrSerProSerSer.....ProAlaIle..... 695
966 ATTGTGAGGCGTCCAAATATAGTCGCG.....CCTT 932
696 .....ProAlaIleSerIleSerIleProMetPheIleProI 708
931 GTTAATCGCGCGCTTCATATACGAA..... 906

```

```

708 IePheThrAlaProProIleSerGlyGlyIleGlyProIleProIleProIle 724
      ||||| ::||| ::|||
905 CCGGAATACAGCGGTTGTGTCAACCAATTCGCGCGCAGCAATTTG 856
      ||| ::||| ::||| ::|||
725 ProSerValThrAlaThrAlaProSerSerSerSerIleProThrThr 741
855 CGATCTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 806
      ::||| ::||| ::|||
741 rSerThrThrAlaPro..... 746
805 GAGAACCACTAAGGCAATACAGCGGCTGGTGTTCAGCGCGCTGTGCA 756
746 ..... 746
755 AACAAAGCGCAATGTAATTAATTCATTTGATTAATGATGTCACACGCT 706
746 ..... 746
705 TTTATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 746
      ||| ::||| ::|||
747 ThrPheGlnProValPheSer..... 753
655 CGGCGAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 606
      ::||| ::||| ::|||
754 .....SerMetGlyProProAlaSer...ValProLeuProAlaProPhe 767
605 TCAGACGCGACGTCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 556
      ::||| ::||| ::|||
768 PheIleGlnThrThrThrProAlaThrAlaProThr..... 779
555 GGTCAAGCGCTCAATACCAACAGCGCGCGCGCGCGCGCGCGCGCG 506
80 ..... 783
505 CTTTGAATTAATGACGTAGGTCGCGCGCGCGCGCGCGCGCGCGCG 456
      ||| ::||| ::||| ::|||
783 roleuPheThrGlyLeuAlaSerIleThrSerAlaValAlaProIleThr 799
455 TTGCAAGAGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 412
      ::||| ::||| ::||| ::|||
800 SerAlaSerProSerThrThrAspSerAlaSerIleProAlaPheGlyPheI 816
411 GCTGAAGGAGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 362
816 YIleAsnSer...ValSerSerSerValSerThrThrThrSerThr 832
361 TCGCGCGCACTTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 318
      ::||| ::||| ::||| ::|||
832 IeThrAlaIleSerIleProPheLeuPheGlyAlaProGlnAlaSerAla 848
317 GGTAGCGCTCAACTGATTTGCTGTTG.....CCTT 283
      ||| ::||| ::||| ::|||
849 AlaSerPheThrProAlaMetGlySerIlePheGlnPheGlyIleProP 865
282 AACGCAATCAAGACTGACTGAATAGCGCGCTTTCCGCGCGGTGAATG 223
865 oAlaLeuProThrThrThrThrValThrThrPheSerGlnSer..... 879
232 CGGCGATTTGCTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 195
      ||||| ::||| ::||| ::|||
880 .....LeuProThrAlaValAlaProThrAlaThrSerSerSerAlaIle 893
194 .....CCGGAATCTTTGCTCTTCAACAGCACTTGCGCTTTTTCAG 151
894 AspPheSerGlyPheGlySerThrIleAlaThrSerAlaProAlaThrS 910
150 GGCATCGCGCTTCTGACTTTCATCGAGGCGCGCGCGCGCGCGCGCG 101
      ::||| ::||| ::||| ::|||
910 rSerGlnProThrIleThrThrPheSerAsnThrSerThrProThrPheAsnI 927
100 CGCAGCAAGCGCGCTTCTGTAATGCGCGCGCGCGCGCGCGCGCG 51
      ||| ::||| ::||| ::|||

```

927 leproheglyseralalysSerProleuProSerTy-ProGlyAla 943

50 TCCGCTCCCGCATGGGC 30

944 AsnProGlnProAlaPhegly 950

seq\_name: SwissProt\_40:MURA\_ECOLI

seq\_documentation\_block:

ID MURA\_ECOLI STANDARD; PRN; 419 AA.  
AC P28909;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE UDP-N-acetylglucosamine 1-carboxyvinyltransferase (EC 2.5.1.7)  
DE (Enolpyruvate transferase) (UDP-N-acetylglucosamine enolpyruvyl transferase) (EPT)  
GN MURA OR MURZ OR B3189.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.  
RC STRAIN=K12 / AB1157;  
RX MEDLINE=92380955; PubMed=1512209;  
RA Margardt J.L., Siegele D.A., Kolter R., Walsh C.T.;  
RT "Cloning and sequencing of Escherichia coli murZ and purification of its product, a UDP-N-acetylglucosamine enolpyruvyl transferase.";  
RL J. Bacteriol. 174:5748-5752(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1233-1238(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O26 / NGY47;  
RX MEDLINE=99216887; PubMed=10103182;  
RA Horii T., Kimura T., Sato K., Shibayama K., Ohta M.;  
RT "Emergence of fosfomycin-resistant isolates of Shiga-like toxin-producing Escherichia coli O26.";  
RL Antimicrob. Agents Chemother. 43:789-793(1999).  
RN [4]  
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).  
RX MEDLINE=97148340; PubMed=8994972;  
RA Skarzynski T., Mistry A., Monacott A., Hutchinson S.E., Kelly V.A., Duncan K.;  
RT "Structure of UDP-N-acetylglucosamine enolpyruvyl transferase, an enzyme essential for the synthesis of bacterial peptidoglycan, complexed with substrate UDP-N-acetylglucosamine and the drug fosfomycin.";  
RL Structure 4:1465-1474(1996).  
RN [5]  
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).  
RX MEDLINE=96151140; PubMed=9485407;  
RA Skarzynski T., Kim D.H., Lees W.J., Walsh C.T., Duncan K.;  
RT "Stereochemical course of enzymatic enolpyruvyl transfer and catalytic conformation of the active site revealed by the crystal structure of the fluorinated analogue of the reaction tetrahedral intermediate bound to the active site of the C15A mutant of MurA.";  
RL Biochemistry 37:2572-2577(1998).  
CC -1- FUNCTION: CELL WALL FORMATION. ADDS ENOLPYRUVYL TO UDP-N-ACETYLGLUCOSAMINE. TARGET FOR THE ANTIBIOTIC PHOSPHOMYCIN.  
CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + UDP-N-acetyl-D-glucosamine = phosphate + UDP-N-acetyl-3-O-(1-carboxyvinyl)-D-glucosamine.  
CC -1- PATHWAY: FIRST COMMITTED STEP IN PEPTIDOGLYCAN BIOSYNTHESIS.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
CC -1- SIMILARITY: BELONGS TO THE EPPS SYNTHASE FAMILY. MURA SUBFAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to [license@isb.stb.ch](mailto:license@isb.stb.ch)).  
CC  
CC EMBL: M92358; AAA24187.1; -  
CC EMBL: U18997; AAA57990.1; -  
CC EMBL: AE000399; AAC76221.1; -  
CC EMBL: AB028039; AAC78107.1; -  
CC PIR: A44917; A44917  
CC PDB: 1UAE; 04-SEP-97.  
CC PDB: 1A2N; 29-APR-98.  
CC Ecogene: EG11358; murA.  
CC InterPro: IPR001986; EPPS\_synthase.  
CC Pfam: PF00275; EPPS\_synthase; 1.  
CC ProDom: PD001867; EPPS\_synthase; 1.  
CC Peptidoglycan synthesis: Cell wall; Cell division; Transferase;  
CC 3D-structure; Complete proteome.  
CC ACT\_SITE 115 115  
CC FT SEQUENCE 419 AA; 44817 MW; 6575A842255E53F7 CRC64;  
SQ

alignment\_scores: Quality: 102.00 Length: 350  
Ratio: 0.662 Gaps: 15  
Percent Similarity: 44.000 Percent Identity: 20.286

alignment\_block: US-09-303-518D-125 x MURA\_ECOLI

Align seg 1/1 to: MURA\_ECOLI from: 1 to: 419

67 GGGCCGGCCATTAC.....GAAGTCGGTCTGGCGGA 101  
||||| :  
98 GlyProLeuValAlaArgPheGlyGlnGlnValSerLeuProGly 114  
102 AGAATATGCGCGTATGCGCC..... 123  
||| |||||  
114 YcstThrIleGlyAlaArgProValAspLeuHisIleSerGlyLeuGln 131  
124 .....TCGATGAAGTCAAGGAGC..... 144  
:  
131 InLeuGlyAlaThrIleLeuGlnLeuGlnIleGlyValIleValSerVal 147  
145 GATGCCGCAAAAAGGCAAGTCGCTGTGAAGCAAAAAGATCCGGG 194  
||||| :  
148 AspGlyArgLeuLysGlyAlaHisIleValMetAspLysValSerValG 164  
195 CGTGGTGTACTCGCGCGCTTCAGGCAAAATCGCCGATTCACCGTG 244  
:  
164 YAlaThrValThrIleMetCysAlaAlaThrLeuAla..... 176  
245 GCGAAAAGCCGCTACTCAGTCAGTCGATTCGCGTGAAGCAACGAC 294  
||||| :  
177 .....GluGlyThrThr 180  
295 GAATCGAGTTGAACGCTACGACCTGAAGCGCTG..... 330  
||||| :  
181 IleIleGluAsnAlaAlaArgGlnProGlnIleValAspThrAlaAsnPh 197  
331 .....GCAACTTAAGCGCGGAGAGAGAGTCCGCGCAAC 364  
:  
197 eLeuIleThrLeuGlyAlaLysIleSerGlyGlnGlyThrAspArgTleV 214  
365 TGAATCAA.....TCGGTGTGGAGTCCGCTCGC 396  
:  
214 aIleIleGlyValAlaGlnArgLeuGlnGlyGlyValThrArgValLeuPro 230

```

397 ACCGCTCGCTTCAGCAAAATCTGCGTGCATGCGAGCGTTCGGCAT 446
    |||
231 Asparg.....IleuThrIleuVal 240
447 CTGCTCATGCGATGACACCAATCCGTGGCTGCGACCGCATGAGTCA 496
    |||
240 LAlaVal.....AlaIleSerArgGlyLysIle 250
497 TTATCAAGAAGCCGCGAGATTTCAAGCGCCGCTGGTATGAGC 546
    |||
250 IecysArgAsnAlaGlnProAspThrLeuAspAla.....ValLeuAla 264
547 CGTTGACCCGACCAAAATCCATGTTTGTAGCAGCTGCGCAGACG 596
    |||
265 LysLeuArgAsp.....AlaGlyAlaAspIle 273
597 G.....CCGCTGAAA 607
273 egluValGlyGluAspTrpIleSerLeuAspMetHisGlyLysArgPro 290
608 ATGCTGCCAATCGAAACACATGAAATTCGCGCGCCGCTGCTGCGGT 657
    |||
290 ysaAlaValAsnValArgThr.....AlaProHisProAlaPhe 302
658 TTGAGTGGCAGCAGCATTCATTTC.....ATCGAGC 689
    |||
303 ProThrAspMetGlnAlaGlnPheThrLeuAsnLeuValAlaGluG 319
690 GGTGCGC...GCGAATAAAACCGTGGGAGCATCAATATATCAAGTGA 736
    |||
319 yThrGlyPheIleThrGluThrValPheGluAsnArgPheMetHisVal 336
737 TTCACATGGCGCTGTTGTTGCAACAGCGCGCTGACACCGAGCGCTG 786
    |||
336 rosluLeuSerArgMetGlyAlaHisAlaGlnIleGluSerAsnThrAl 352
787 ATGCGCTGAGTGGTTCCTCAAGTCAACAACCGCGCTTGGCGTACCG 836
    |||
353 IleCysHisGly..... 356
837 TTGGGTGGGAAAGTATCGCAATTAATGCGGCGAATGGTGCACAG 886
    |||
357 .....ValGluLysLeuSerGlyAlaGlnValMetAlaThrAl 369
887 ACAACCGCGTGAATTCGCGTTCGATGACAGCGCGCATTAACAAGC 936
    |||
369 spleuArgAlaSerAlaSerLeuValLeuAlaGlyCysIleAlaGluGly 385
seq_name: SwissProt_40:ODP2_ZYMO
seq_documentation_block:
ID ODP2_ZYMO STANDARD: PRT: 440 AA.
AC 066119:
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase
DE complex (EC 2.3.1.12) (E2).
GN PGHC OR PDHB.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
OC Zymomonas.
OX NCBI_TaxID=542:
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29191 / ZM6.
RX MEDLINE=98175679; PubMed=9515924;
RA Neveling U., Klaesen R., Bringer-Meyer S., Sahm H.;
RT "Purification of the pyruvate dehydrogenase multienzyme complex of
RT Zymomonas mobilis and identification and sequence analysis of the
RT corresponding genes.";
RL J. Bacteriol. 180:1540-1548(1998).

```

```

CC -1- FUNCTION: THE PYRUVATE DEHYDROGENASE COMPLEX CATALYZES THE OVERALL
CC CONVERSION OF PYRUVATE TO ACETYL-COA & CO(2). IT CONTAINS MULTIPLE
CC COPIES OF THREE ENZYMATIC COMPONENTS: PYRUVATE DEHYDROGENASE (E1),
CC DIHYDROLIPOAMIDE ACETYLTRANSFERASE (E2) & LIPOAMIDE DEHYDROGENASE
CC (E3) (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + dihydrolipoamide -> CoA + S-
CC acetyldihydrolipoamide.
CC -1- COFACTOR: THE E2 COMPONENT CONTAINS ONE COVALENTLY-BOUND LIPOYL
CC COFACTOR (BY SIMILARITY).
CC -1- SUBUNIT: FORMS A 24-POLYPEPTIDE STRUCTURAL CORE WITH OCTAHEDRAL
CC SYMMETRY (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE 2-OXOACID DEHYDROGENASE FAMILY.
CC -1- SIMILARITY: CONTAINS 1 LIPOYL-BINDING DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X93605; CAA63808.1; -.
CC HSSP: P10515; IRYC.
CC DR InterPro: IPR001078; 2Oxoacid_dh.
CC DR InterPro: IPR000089; Biotin_lipoyl.
CC DR InterPro: IPR003016; Lipoyl.
CC DR Pfam: PF00198; 2-oxoacid_dh; 1.
CC DR Pfam: PF00364; biotin_lipoyl; 1.
CC DR Pfam: PF02817; e3_binding; 1.
CC DR ProDom: PD001115; 2Oxoacid_dh; 1.
CC DR PROSITE: PS00189; LIPOYL; FALSE-NEG.
CC KW Glycoyls; Transferase; Acyltransferase; Lipoyl.
CC FT BINDING 43 43 LIPOYL (POTENTIAL).
CC FT ACT SITE 412 412 POTENTIAL.
CC SQ SEQUENCE 440 AA; 46778 MW; 173A98739ACA9A40 CRC64;

```

alignment\_scores:

|  | Quality:            | 102.00 | Length:           | 415    |
|--|---------------------|--------|-------------------|--------|
|  | Ratio:              | 0.510  | Gaps:             | 15     |
|  | Percent Similarity: | 48.193 | Percent Identity: | 21.205 |

alignment\_block:

US-09-303-518d-125 x ODP2\_ZYMO ..

Align seg 1/1 to: ODP2\_ZYMO from: 1 to: 440

```

70 CCGGCGATTACGGAAGTCCGCTTCTGGCGAAGATATGCGCGTATGCG 119
    |||
12 ProThrMetThrGlu.....GlyThrIle 19
120 CCCCTCGATGAAGTCAAGAGGCGATGCCGTCAAAAAGGCCAGTGC 169
    |||
19 uAlaIleuValValLysGluGlyAspAlaValValLysAlaGlyAspIle 36
170 TGTTTGAAGACAAAAGATCCGGCGGTG...TTTACTGCGCGCGGT 216
    |||
36 euAlaGluIleGluThrAspLysAlaIleMetGluPheGluThrValAsp 52
217 TCAGCAAAATCGCGCGGAT.....CACCGGCGAAGAGCGCGTACT 260
    |||
53 AlaGlyIleIleAlaLysIleLeuValProGluGlySerGluAsnIleAl 69
261 TCATCATGCTGATTCGCTTGAAGCAACGCAAGATCGATTGTGAAC 310
    |||
69 aValGlyValIleAlaValMetAlaGluAlaGlyGluAspValSerG 86
311 GCTACGCGACCTGAAGCGCTGCAAACTTAAGCGGCAAGAACTGGCGCC 360
    |||
86 InValAlaValaSerAlaSerSerGlnIleSerGluProSerGluLysAla 102

```

```

361 AACCTGATCCAAATCCGGTTTGTGACTGCGTCCGACCCGTCGTTGAC 410
      |||
103 AspValAlaGln.....LysGluThrAlaAspSe 112
411 CAAATTCCTCCGTCGATGCCGACCGTCGCCATCTTCGTCATGGA 460
      |||
112 TGTuThrIleSerIleAspIleSerLeuAspLysAlaIleSerAsnIaG 129
461 TGGACACCAATCCGTCGTCGCCGACCCATGATTCATTAAGAACGCC 510
      |||
129 IyYrGlyAsnLys...ThrGluAsnMetThrAlaSerTyrGlnIlu... 143
511 GCCGAGGATTTCAACCGCGCCTGTGTGATTGACCCGTTGACGCAAG 560
      |||
144 .....LysAlaGlyArgIleLysAlaSerProLeuAlaLysAr 156
561 CAAATTCATGTTGTAAAGCAGCTGCGCAGACGTCGCGTGTGAATG 610
      |||
156 9.....LeuAlaLysLysAsnH 162
611 CTGCCAATCGAACAACATGAAATTCGGCGCCGATCTCCGCTTG 660
      |||
179 LysAlaAspIleGluAlaPheValThrGlyAlaAsnGlnAlaSerSers 195
661 AGTGGACGACATTCATTCATC.....GACCGGTCGCGCGCA 701
      |||
702 TAAACCGTGTGACCATCAATATCAAGATGTAATACCATGGCCGT 751
      |||
195 nProSerValSerThrProGlnValSerGlyLysIleThrHisAspThr 212
752 TGTGTTCACAGCGCCGTCTG...AACCGGCGCGGTGATGCC..... 792
      |||
212 roHisAsnSerIleLysLeuSerHisMetArgValIleAlaArg 228
793 CTAGGTGTTCTCAAGTCAACAAACCGCGCTTCGTCGACGTT..... 837
      |||
229 LeuThrGluSerLysGlnAsnIleProHisIleTyrLeuThrValAspVa 245
838 .....TTGG 841
245 IGlMetAspAlaLeuLysLeuArgSerGluLeuAsnGluSerLeu 262
842 GTGCGAAGTATCGCAATCTACTGCGCGCATTTGTT..... 879
      |||
262 IavaGlnAsnIleLysIleSerValAsnMetLeuIleLysAlaGln 278
880 .....GACACAGA 887
279 AlaLeuAlaLeuLysAlaThrProAsnValAsnValAlaPheAspGlyAs 295
888 CAACCGCGCTG.....ATTTCGGTTCGGTAT 913
      |||
295 pGlnMetLeuGlnPheSerGlnAlaAspIleSerValAlaValSerValG 312
914 TGAACGGCGCATTCACACAGCGCGACATTTTGGGACGCTACCAC 963
      |||
312 IuGlyIleuThrProIleLeuLysGlnAlaAspThrLysSerLeu 328
964 AATTCAGATTTCCGTCATGAGAGAGCGCAGCAAGGCGTCGCGCTG 1013
      |||
329 SerAlaLeuSerVal.....GluMetLysGluLeuIleAlaAr 341
1014 GGTTCGCGCG.....CAGCGCGCAATCTCATCATCAGCGGTACAA 1054
      |||
341 gAlaIArgGluGlyArgLeuGlnProGlnGluTyrGlnIlyGlyThrSers 358
1055 CCCTGGCGCATTTCTGAAAAACAACCTTCAGTTCAACACAGCCGCTC 1104
      |||
358 erIleSerAsnMetGlyMetPheGlyIleLysGlnPheAsnAlaValIle 374
1105 AACGGCGGCGACCGGCATGTGTGCGCATGTGTACTTACGAGCGC 1149

```

```

seq_name: SwissProt_40:PUR2_CHICK
375 AsnProGlnAlaSerIleLeuAlaIleLysGlyArg 389
      |||
      ::::::::::: |||||::: |||||
seq_documentation_block:
ID PUR2_CHICK STANDARD; PRT; 1003 AA.
AC P21872;
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-MAY-1991 (Rel. 39, Last annotation update)
DE Trifunctional purine biosynthetic protein adenosine-3 [includes:
DE phosphoribosylamine--glycine ligase (EC 6.3.4.13) (GARS) (glycinamide
DE ribonucleotide synthetase) (phosphoribosylglycinamide synthetase)
DE phosphoribosylformylglycinamide cyclase (EC 6.3.3.1) (AIRS)
DE (phosphoribosyl-aminimidazole synthetase) (AIR synthase);
DE phosphoribosylglycinamide formyltransferase (EC 2.1.2.2) (GART) (GAR
DE transformylase) (5'-phosphoribosylglycinamide transformylase)].
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=91067455; PubMed=2147474;
RA Aimi J., Olu H., Williams J., Zalkin H., Dixon J.E.;
RT "de novo purine nucleotide biosynthesis: cloning of human and avian
RT cDNAs encoding the trifunctional glycinamide ribonucleotide
RT synthetase-aminimidazole ribonucleotide synthetase-glycinamide
RT ribonucleotide transformylase by functional complementation in E.
RT coli.";
RL Nucleic Acids Res. 18:6665-6672(1990).
CC -1- CATALYTIC ACTIVITY: ATP + 5-phosphoribosylamine + glycine = ADP +
CC phosphate + 5'-phosphoribosylglycinamide.
CC -1- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + 5'-
CC phosphoribosylglycinamide = tetrahydrofolate + 5'-
CC phosphoribosylglycinamide.
CC -1- CATALYTIC ACTIVITY: ATP + 5'-phosphoribosylformylglycinamide =
CC ADP + phosphate + 5'-phosphoribosyl-5-aminimidazole.
CC -1- PATHWAY: SECOND, THIRD AND FIFTH STEPS IN DE NOVO PURINE
CC BIOSYNTHESIS.
CC -1- ALTERNATIVE PRODUCTS: THE GARS DOMAIN CAN BE SYNTHESIZED AS A
CC SEPARATE PROTEIN BY ALTERNATIVE SPLICING OF THE GENE.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION: BELONGS TO THE GARS FAMILY.
CC -1- SIMILARITY: TO OTHER AIRS AND GART FROM BACTERIA AND EUKARYOTES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X54200; CAA38120.1; -.
DR PIR: S12617; ATCHPR.
DR HSSP: P15640; IGSO.
DR InterPro: IPR000728; AIRS_related.
DR InterPro: IPR000115; GARS.
DR InterPro: IPR001555; GART.
DR InterPro: IPR002376; formyl_transf.
DR Pfam: PF00586; AIRS; 1.
DR Pfam: PF02769; AIRS_C; 1.
DR Pfam: PF00351; formyl_transf; 1.
DR Pfam: PF01071; GARS; 1.
DR Pfam: PF02842; GARS_B; 1.
DR Pfam: PF02843; GARS_C; 1.
DR Pfam: PF02844; GARS_N; 1.
DR PROSITE: PS00184; GARS; 1.
DR PROSITE: PS00373; GART; 1.

```



KW Multifunctional enzyme; Purine biosynthesis; Ligase; Transferase;  
 Alternative splicing.  
 FT DOMAIN 1 433 GARS.  
 FT DOMAIN 434 805 AIRS.  
 FT DOMAIN 806 1003 GART.  
 FT ACT\_SITE 947 947 BY SIMILARITY.  
 FT VARSPEC 434 1003 MISSING (IN ISOFORM GARS-ONLY).  
 SQ SEQUENCE 1003 AA; 106544 MW; A0C66BA0EBF791DE CRC64;

alignment\_scores:  
 Quality: 102.00 Length: 490  
 Ratio: 0.479 Gaps: 25  
 Percent Similarity: 43.469 Percent Identity: 21.429

alignment\_block:

US-09-303-518D-125 x PUR2\_CHICK ..

Align seg 1/1 to: PUR2\_CHICK from: 1 to: 1003

```

7 AAAATCAAAAAGGTCTAACTGACCCATCGCGGCGAGACCGAGCAACG 56
  ::::::::::: ||| ||| |||
241 GlnIleSerLysAspLeuLeuLysIleArgGluThrValLeuGlnIly 257
  ::::::::::: ||| ||| |||
57 CGTTACGAGCGCGCGCCATTACGCAAGTCGGTCTTGCGCAAGAT 106
  ::::::::::: ||| ||| |||
257 sTrhLeuAspGlyMetArgLysGlnIlyLeuProLysGlyValLeu 274
  ::::::::::: ||| ||| |||
107 ATCCGGTATGCGCGCCCTGATGAAGTCAAGAGCGCATCGCTCAAA 156
  ::::::::::: ||| ||| |||
274 yrnIagIlyLeu.....MetLeuThrLysAspLysProLysValLeu 287
  ::::::::::: ||| ||| |||
157 AAGGCCAAGTGTCTTTGAAGAC.....::: 180
  ::::::::::: ||| ||| |||
288 GluPheAsnLysArgPheGlyAspProGluLysGlnValIleLeuPro 304
  ::::::::::: ||| ||| |||
181 ....AAAAGAAATCCGCGGTGTTTACGCGCGCGCTCAGCAAAA 226
  ::::::::::: ||| ||| |||
304 uLeuLysSerAspLeuLysGlnValMetGlnIleValIleAsnLysIly 321
  ::::::::::: ||| ||| |||
227 TCGCCCGCATTCACCGT.....GGCGAAAGCGGTACTTCAGTCACTG 270
  ::::::::::: ||| ||| |||
321 euSerSerMetProIleThrLysGlnValSerAlaIleValThrVal 337
  ::::::::::: ||| ||| |||
271 GTGATTGCGGTGAAGGACAGCAATTCAGTTTGACCGCTACGACG 320
  ::::::::::: ||| ||| |||
338 ValMetAlaSerGluIly.....TyrProGlyThrTyrPr 349
  ::::::::::: ||| ||| |||
321 TGAACGCTG.....GCAACTTAAGC...GGCGAAGAGTGGCGCGCA 361
  ::::::::::: ||| ||| |||
349 oLysGlyLeuGlnIleThrGlyLeuSerLysAlaLysGlnLeuGlyLeu 366
  ::::::::::: ||| ||| |||
362 ACCGTATCCCAATCCGCTTTGTGACTGCGCTGCGCACCCGCTCGTCA 411
  ::::::::::: ||| ||| |||
366 lUValPheHisIleGly.....ThrIleLeuLys..... 375
  ::::::::::: ||| ||| |||
412 AAAATTCCTGCGCGATGCGAGCCGCTGCGCATCTTCGTCATGCGAT 461
  ::::::::::: ||| ||| |||
376 .....AspGlyLysValValThrAsnLysGlyValLe 387
  ::::::::::: ||| ||| |||
462 GGACACCAATCCGCTGCGCTGCGCACCTACGCTATTCATCAAGAA... 507
  ::::::::::: ||| ||| |||
387 uThrValThrAlaIleLysGlnAspLeuMetThrAlaLeuGlnIleAla 404
  ::::::::::: ||| ||| |||
508 .....GCCCGAGGATTCAAA..... 525
  ::::::::::: ||| ||| |||
404 snLysGlyValAlaIleAlaIleAsnPheLysLysIleThrArgLysAsp 420
  ::::::::::: ||| ||| |||
526 .....CGCGCGCTGTGTGATTCAGCGCT.....TTGACCGCA 557
  ::::::::::: ||| ||| |||
421 lIleGlyThrArgAlaIleAlaPheLeuSerGlnSerArgGlyLeuThr 437
  ::::::::::: ||| ||| |||
558 ACGCAAAATCCATGTTGTGAAGCAGCTGGC..... 588
  ::::::::::: ||| ||| |||

```

```

437 rLysAsnSerLysValAspIleAlaIleGlyAsnIleLeuValGlnLysI 454
  ::::::::::: ||| ||| ||| ||| |||
589 .....GCAGACGTGCGCTGTGAATAATCTGCAACATCGAACA 627
  ::::::::::: ||| ||| ||| ||| |||
454 lLysProLeuAlaIleAlaIleThrSerArgSerGlyCysAsnAlaGlu 470
  ::::::::::: ||| ||| ||| ||| |||
628 CATGAATTCGCGCGC.....CCGCATCC 650
  ::::::::::: ||| ||| ||| ||| |||
471 GlyLysPheAlaGlyLeuPheAspLeuLysAlaIleGlyTyrLysAsp 487
  ::::::::::: ||| ||| ||| ||| |||
651 TCGCGGTTCAGTGCACG..... 669
  ::::::::::: ||| ||| ||| ||| |||
487 oIleLeuValSerGlyThrAspGlyValGlyThrLysLeuLysIleAla 504
  ::::::::::: ||| ||| ||| ||| |||
670 .....CACATTCATTCATGACCGCGCTCGCGCGAATAA 705
  ::::::::::: ||| ||| ||| ||| |||
504 lValLysLysLysHis.....AspThrIleGlyLysAsp 516
  ::::::::::: ||| ||| ||| ||| |||
706 ACCGTGACCATCATTCATTCAGATGATTAATTCATTCGCGCGCT... 750
  ::::::::::: ||| ||| ||| ||| |||
517 ValAlaMetCysValAsn.....AspIleLeuAlaGlnGlyAlaGluPr 531
  ::::::::::: ||| ||| ||| ||| |||
751 .....TTGTTTGAACAGCGCGCTGGAACACGAG... 780
  ::::::::::: ||| ||| ||| ||| |||
531 oLeuPhePheLeuAspTyrPheAlaCysGlyLysLeuAspValGluVal 548
  ::::::::::: ||| ||| ||| ||| |||
781 .....CGCGTATTCG... 792
  ::::::::::: ||| ||| ||| ||| |||
548 lAsnIleValIleAlaGlyIleAlaGlnAlaCysGlnLysAlaGlyCys 564
  ::::::::::: ||| ||| ||| ||| |||
793 .....CTAGGTGTTCTCAAGTCAACAAACCGCGCTCTTG... 828
  ::::::::::: ||| ||| ||| ||| |||
829 ....CGTACGTTTGGCGGCAAGTATCCCAATTCCTCGCGCGAT 874
  ::::::::::: ||| ||| ||| ||| |||
581 yGluThrAspLeuAlaGlyPheAlaValAlaValAlaGlnArgGlyGln 598
  ::::::::::: ||| ||| ||| ||| |||
875 TGGTTGACACAGACACCGCGATTCGCGTTCGATTCGATTCGAGCGCG 924
  ::::::::::: ||| ||| ||| ||| |||
598 eLeuProGlnLeuGlnArgIleAlaAspGlyAspValAlaIleGlyVal 614
  ::::::::::: ||| ||| ||| ||| |||
925 ATTACCAAGCGCGCGCATTCATTCGAGCGCTACCAATCATGATTC 974
  ::::::::::: ||| ||| ||| ||| |||
615 AlaSerGlyValHisSer..... 621
  ::::::::::: ||| ||| ||| ||| |||
975 CGTTATCGAAGAGCGCGACCAAGAGCTGTCGCTGGGTTGCGCGCG 1024
  ::::::::::: ||| ||| ||| ||| |||
621 ..... 621
  ::::::::::: ||| ||| ||| ||| |||
1025 AAGCGCAATATCTCATCAGCGTACAAACCTCGGCCATTCCTGAAA 1074
  ::::::::::: ||| ||| ||| ||| |||
622 .....AsnGlyTyrSerLeuValArgIleValGln.....Lys 633
  ::::::::::: ||| ||| ||| ||| |||
1075 AACAAATCT...TTCAGTTCACACAGCGCTCAAGCGC..... 1110
  ::::::::::: ||| ||| ||| ||| |||
634 SerSerLeuAspPheSerSerGlnValGlyValSerGlyAspGlnThr 650
  ::::::::::: ||| ||| ||| ||| |||
1111 .GGGACCGCGCGCATGTGCGCATGTTGTTACTTCAAGCGCGCTGATCCCT 1159
  ::::::::::: ||| ||| ||| ||| |||
650 uGlyAspLeuLeuThrProThrLysIleTyrSerLysThr..... 664
  ::::::::::: ||| ||| ||| ||| |||
1160 TGGATATCTGCGCACCGCTG 1179
  ::::::::::: ||| ||| ||| ||| |||
665 .....LeuLeuProValLeu 669
  ::::::::::: ||| ||| ||| ||| |||

```

seq\_name: SwissProt\_40:HFCL\_MESAU  
 seq\_documentation\_block:  
 ID HFCL\_MESAU STANDARD; PRT; 2090 AA.  
 AC P51611;  
 DT 01-OCT-1996 (Rel. 34, Created)

01-OCT-1996 (Rel. 34, last sequence update)  
 16-OCT-2001 (Rel. 40, last annotation update)  
 DE Host cell factor C1 (HCF) (VP16 accessory protein) (HFC1) (VCAF)  
 DE (HCF).  
 GN HFC1.  
 OS Mesocricetus auratus (Golden hamster).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 CC Mesocricetus.  
 CC NCBI\_TaxID=10036.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Goto H., Nishitani H., Umene K.I., Nakabeppu Y., Nishimoto T.;  
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: UPON LYTIC INFECTION OF PERMISSIVE CELLS, THE HSV  
 CC TRANSCRIPTIVATOR PROTEIN VP16 ASSOCIATES WITH HCF. BINDING TO HCF  
 CC ACTIVATES VP16 FOR ASSOCIATION WITH THE OCTAMER MOTIF-BINDING  
 CC PROTEIN OCT-1, TO FORM A MULTIPROTEIN-DNA COMPLEX RESPONSIBLE FOR  
 CC ACTIVATING TRANSCRIPTION OF THE HSV IMMEDIATE EARLY GENES (BY  
 CC SIMILARITY).  
 CC -1- SUBUNIT: COMPOSED PREDOMINANTLY OF SIX POLYPEPTIDES RANGING FROM  
 CC 110 TO 150 KDA AND A MINOR 300 KDA POLYPEPTIDE. THE MAJORITY OF N-  
 CC AND C-TERMINAL CLEAVAGE PRODUCTS REMAIN TIGHTLY, ALBEIT  
 CC NONCOVALENTLY, ASSOCIATED (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- DOMAIN: THE HCF REPEAT IS A HIGHLY SPECIFIC PROTEOLYTIC CLEAVAGE  
 CC SIGNAL (BY SIMILARITY).  
 CC -1- PTM: EXISTS AS A FAMILY OF POLYPEPTIDES THAT ARISE THROUGH  
 CC CLEAVAGE AT ONE OR MORE CENTRALLY LOCATED SITES. CLEAVAGE OCCURS  
 CC AT A DEFINED SITE, PPCE/THEET, WITHIN THE HCF REPEAT (BY  
 CC SIMILARITY).  
 CC -1- SIMILARITY: CONTAINS 5 KELCH REPEATS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: DA5419; BAA08258.1; -  
 CC HSP: P02751; 1FNA.  
 CC InterPro: IPR003961; FN.III.  
 CC InterPro: IPR001798; Kelch.  
 CC Pfam: PF00041; fn3; 2.  
 CC Pfam: PF01344; Kelch; 4.  
 CC SMART: SM00060; FN3; 2.  
 CC KW Nuclear protein; Repeat.  
 FT REPEAT 44 89  
 FT REPEAT 93 140  
 FT REPEAT 148 194  
 FT REPEAT 217 265  
 FT REPEAT 266 313  
 FT REPEAT 313 361  
 FT DOMAIN 1010 1448  
 FT REPEAT 1010 1035  
 FT REPEAT 1072 1097  
 FT REPEAT 1101 1126  
 FT REPEAT 1157 1182  
 FT REPEAT 1295 1320  
 FT REPEAT 1323 1348  
 FT REPEAT 1358 1383  
 FT REPEAT 1423 1448  
 FT SEQUENCE 2090 AA; 214942 MW; E495E8B1F2385E17 CRC64;

alignment\_scores:  
 Quality: 102.00 Length: 462  
 Ratio: 0.502 Gaps: 23  
 Percent Similarity: 43.939 Percent Identity: 21.861

alignment\_block:  
 US-09-303-518D-125/rev x HFC1\_MESAU ..

Align seg 1/1 to: HFC1\_MESAU from: 1 to: 2090  
 1112 CCGCCGTTGAGCGCTGTGTTG.....ACTTGAAGAGTTGTGTTT 1072  
 366 ProProAlaArgValGlnLeuValAlaGlnLeuValAlaSerLeuGlnVal 382  
 1071 CAGGAATAGGCGGAGGTTGAGCGGTGATGAGTATG..... 1032  
 382 L...SerTptGlyAlaValAlaThrAlaAspSerTyrLeuLeuGlnLeu 398  
 1031 .....TCGGCTGCGGCGCACCCAGCCGACAC 1002  
 398 InLysTyrAspIleProAlaThrAlaAlaThrAlaThrSerProThr... 413  
 1001 TCTTTGCTGGCGCTTCTTCATACGAATTCGATGTGTAGCTGC 952  
 414 .....ProAsnProAlaProSerValProAla.....AsnPr 424  
 951 CAAATATGTCGCGCCTTGATGCGCGC..... 918  
 424 oProLysSerProAlaProAlaAlaAlaProAlaValGlnProLeuT 441  
 917 .....TTCAATACGACCGGAATCAGCGGTGTGTGTGCA 879  
 441 hrcGlnValGlyLeuThrLeuValProGlnAlaAlaAlaAlaProProSer 457  
 878 ACCAATTCG.....CCCGCAGTAATTCGATCTTGCCACC 841  
 458 ThrThrThrIleGlnValLeuProThrValProGlySerSerIleSerVa 474  
 840 CAAACGGTAGCGCAGGCGCGT.....TTGTGACTTGACAGAC 800  
 474 LProThrAlaAlaArgAlaGlnGlyValProAlaValLeuLysValThrG 491  
 799 CACTTAGCGCAATCAGCGCTCG.....GTGTGACAGCGCGT... 759  
 491 LProGlnAlaThrThrGlyThrProLeuValThrMetArgProAlaGly 507  
 758 ...GCAACAAACGCGCAATGTGTAATTCATCTGTAATGANGTCCA 712  
 508 GlnAlaGlyLysAlaProValThrValThrSer.....LeuProAlaSe 522  
 711 CAGCGTTTATTCGCGCGCGCGCTCGATGCAATGATGTCGCGCAC 662  
 522 rValArgMetValValProThrGlnSerAlaGlnGlyThrValIleGly 539  
 661 TCAAACGCGCAGGATGCGG..... 642  
 539 erAsnProGlnMetSerGlyMetAlaAlaLeuAlaAlaAlaAlaAla 555  
 641 .....CCGCCGAATTCATGTGTTGATGTTGCA..... 612  
 556 ThrGlnLysIleProProSerSerAlaProThrValLeuSerValProAl 572  
 611 .....GCATTTGAGACGCGACGTCGCGC 587  
 572 agLysThrThrIleValLysThrValAlaValAlaThrProGlyThrThrL 589  
 586 CAGCTGCTTACAAACATGATTTTGGCTTGGTCAACGCGCTCATAC 537  
 589 euProAlaThrValLysValAlaLaserSerProValMetValSerAsnPro 605  
 536 AACAGCGCGCTTGAATCCTGCGGCGCTTGTGTAATGACCGTAG 487  
 606 AlaThrArgMetLeuLysThrAlaAlaAlaGlnValGlyThrSerValSe 622  
 486 GTCGCGACGC.....ACGC 473  
 622 rSerAlaAlaAsnThrSerThrArgProIleIleThrValIleLysSerG 639  
 472 GATTGCTGCATGCA.....TTGACGAAGATGCGGAC 438

```

639 lYThrValThrValAlaGlnGlnAlaGlnValValThrThrValValGly 655
437 GGCTCGCAGTCAGC..... 423
656 GlyValThrThrThrThrThrThrValLysSerProIleSerValProGly 672
422 ....GCAGAAATTTTGGCTGAC...GCAGGGGTGGCAGCGACGACCA 380
672 yglSerAlaLeuIleSerAsnLeuGlyLysValMetSerValValGln. 688
379 AACCGATTTGATCAGTTGGCGGCGCACTTCTCGCGCTTAAGTTGGC 330
689 .....ThlyProValGlnThrSer 695
329 ACCGCTTCAGGTGCTGACGTTCAACATCGATTCGCGCTTG..... 288
696 AlaValThrGlnAlaSerThrGlyProValThrGlnIleIleGlnThr 712
288 ..... 288
712 lYsGlyProLeuProAlaGlyThrIleLeuLysLeuValThrSerAla 729
287 .....CCTCAACGCAATCAGCACTGACGCTGACGCTTTCG* 246
729 spGlyLysProThrThrIleThrThr.....ThrGlnAlaSer 742
245 CCACGGTGAATCGCGGCAATTTGCTGAAGCGCGCAGTAACACAC 196
743 .....GlyAlaGlySerLysProThrIle 750
195 GCCCGATTCCTTTTCTTCAACAGCACTGGCTTTTTCAGCGCAT 146
750 eLeuGlyIleSerSerValSerProSer.....ThrThrL 762
145 CGCCTTCCTTCACTTCATCAGAGCGGCGCATACCG.....GCATATCT 102
762 yspProGlyThrThrThrIleIleLysThrIleProMetSerAlaIleIle 778
101 TCGGCAAGCAAGCACTGCGATGCGGCGGCGGCGG 66
779 ThrGlnAlaGlyAlaThrGlyValThrSerThrPro 790
seq_name: SwissProt_40:AMPH_CHICK
seq_documentation_block:
ID AMPH_CHICK STANDARD; PRT; 682 AA.
AC P50478;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Amphiphsin.
GN Amph.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TERRA-HYBRID; TISSUE=Forebrain;
RX MEDLINE=92331604; PubMed=1628617;
RA Lichte B., Voh R.W., Meyer H.E., Killmann M.W.;
RT "Amphiphsin, a novel protein associated with synaptic vesicles.";
RL EMBL J. 11:2521-2530(1992).
CC -1- FUNCTION: MAY PARTICIPATE IN MECHANISMS OF REGULATED EXOCYTOSIS IN
CC SYNAPSES AND CERTAIN ENDOCRINE CELL TYPES. MAY CONTROL THE
CC SUBCELLULAR LOCATION: ASSOCIATED WITH THE CYTOPLASMIC SURFACE OF
CC SYNAPTIC VESICLES.
CC -1- TISSUE SPECIFICITY: IS ABUNDANT IN THE FOREBRAIN AND CEREBELLUM.
CC IT IS ALSO FOUND IN THE ADRENAL GLAND, ANTERIOR AND POSTERIOR
CC PITUITARY.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.

```

```

CC -1- SIMILARITY: BELONGS TO THE AMPHIPHYSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X60422; CAA42933.1; -
DR HSP; P29355; ISEM.
DR InterPro; IPR003005; Amphiphsin.
DR InterPro; IPR004148; BAR.
DR InterPro; IPR001452; SH3.
DR Pfam; PF03114; BAR; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PRO1251; AMPHIPHYSIN.
DR PRINTS; PRO0452; SH3DOMAIN.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
DR Cytoskeleton; SH3 domain; Coiled coil.
FT DOMAIN 10 84 COILED COIL (POTENTIAL).
FT DOMAIN 144 191 COILED COIL (POTENTIAL).
FT DOMAIN 609 682 SH3.
SQ SEQUENCE 682 AA; 75205 MW; 61617P494F38EB20 CRC64;

```

```

alignment_scores:
Quality: 101.50 Length: 371
Ratio: 0.573 Gaps: 17
Percent Similarity: 47.709 Percent Identity: 20.216

```

alignment\_block:

US-09-303-518d-125 x AMPH\_CHICK ..

Align seg 1/1 to: AMPH\_CHICK from: 1 to: 682

```

31 CCCATGGCGGGGAGACCGAGCA.....GCCGTTACAGCGCCGCGC 74
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
286 ProAlaProAlaArgProLysSerProThrGlnLeuArgLysGlyProPr 302
75 CATACCGCAAGTCGCGTGTGCGGCAAGAAATATCCGCTATCGCCCT 124
||| :|||:|||||:|||||:|||||:|||||:|||||:
302 oValProProLeuProLysLeu.....ThrPro 312
125 CGATGAAGTCAAGAGAGCGATGCCGTCAAAAAGCGCAAGTGTGTT 174
: :|||:|||||:|||||:|||||:|||||:|||||:
312 hLysGlnLeuGlnGlnGlnGlnIleIleAsn.....LeuPhe 324
175 GAAGACAAAAGAAATCCGGCGCGTGTGTTACTGCGCGGCTTAGCA 224
: :|||:|||||:|||||:|||||:|||||:|||||:
325 AspAspAsnPheValProGlnIleAsnValThrThrProSerGlnAsn 341
225 AATGCCCGGATTCACCGTGGCAAAAGCGGTACTACGACAGTGTGA 274
||| :|||:|||||:|||||:|||||:|||||:|||||:
341 uIlePro.....GluThrLysLysValGlnSerLeuLeu. 352
275 TTGCGGTTGAAGGACACACCAATCGAGTTGAAGCTACGCACTGA 324
: :|||:|||||:|||||:|||||:|||||:|||||:
353 .....AspLeuAspPheAspPropheLysProGln 362
325 GCGCTGCAAACTTAAGCGCGAGAGAGTGGCGGCAACCTGATCCATC 374
: :|||:|||||:|||||:|||||:|||||:|||||:
363 ValValSerThrGlyValThrHisSerProMetSerGlnThrLeuPro 379
375 CGGTTTGGACTCGCGCTGCGC.....ACCGTCGTTACGCAAAATTC 418
|||||:|||||:|||||:|||||:|||||:|||||:
379 pAspLeuThrThrThrThrSerGlnLeuValGlnProAlaSerSerThra 396
419 CT.....GCCGTCAGCGCGAGCGGCTTCCGCACTTCGTCAT 456
||| :|||:|||||:|||||:|||||:|||||:
396 lApheAsnGlyPheAlaGlnAspThrThrAlaPheAlaValGlnSerAsn 412

```

457 GCGATGACACCAATCGCTGGCGCCGACCCCTACGGTCATTATCAAGA 506  
 413 GUAANVALTHGLTHLEUTHR..... 420  
 507 ACCCGCCGAGATTCAACAGCGCGCTTGGTATGACCCGTTGACCG 556  
 421 GUAAGLUGLUGLUALPROLEUGLUGLULYVALGLULUGLTHRPROT 437  
 557 AACGCAAAATCCATGTTTGAAGCA.....GCGCGCGAGACGCGCG 600  
 437 hVALAVALAVALAVALUGLUGLUALAILEUALAAGLUPROASPLUPRO 453  
 601 TCGAAATGCTGCCACATCGAACACATGATTCGCGCGCCGATCC 650  
 454 THGLUGLUALA.....GLUSERTLEGUALAGLYASPLUGLUTH 468  
 651 TCGCGGTTGAGTGCACGACATTCATTTCATCGACCGCGTGGCGGA 700  
 468 rTHGLYILEALAGLULYGLUSERTGLUVALVALSERALALAGLYLY 484  
 701 ATAAACCGTGTGACCATCATATATCAAGATGATTAATTCATGCGCGT 750  
 485 .....AVALAVALAVALAVALASPSERVALVALAVALA 496  
 751 TTGTTTGCACAGCGCGCTGAAACACGCGCGGATTCGCGCGTGG 800  
 497 GLVALAGLUGLUGLUALAVALATGTHGLUGLUGLUALAVALAGLUGL 513  
 801 TTCTCAAGTCAACAAACCGCGCTCTGGTACCGTTTGGGTGGGAAG 850  
 513 Y.....ASPLYSERGLIN..... 517  
 851 TATCGAAATTAATGCGCGGAATGTTGACACAGACACCGCGTGA 900  
 518 .....GLYGLUGLUGLULYASPLASPLASPLASPLASPLASPLAS 528  
 901 TCCGCTGCGATGTAACGCGCGGATTCACACAGCGCGGATTAAT 950  
 529 LYSVALSERISERILEPROSERVALVALILEGLIPROLASERASPLAS 545  
 951 GGA.....CGTACACACATGATTCGTTATGAGAAGGCC 991  
 545 uGLYGLUGLUGLUGLULHSHS.VALLIEMETASNGLUSERTLYSASPLA 561  
 992 GCAGCA.....AAGAGCTGTCGCGCTG 1014  
 562 AAlaAGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGL 578  
 1015 GTTGGCGCGAGCGGACCA.....AATACTCATACGCGCTAC 1052  
 578 YSERGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGLUGL 595  
 1053 AACCTCGGCC 1063  
 595 LIPROLASER 598  
 seq\_name: Swissprot\_40:CSDB\_ECOLI  
 seq\_documentation\_block:  
 ID CSDB\_ECOLI STANDARD: PRT: 406 AA.  
 AC P77444;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Selenocysteine lyase (EC 4.4.1.16) (Selenocysteine reductase)  
 DE (Selenocysteine beta-lyase) (SCU).  
 DE CSDB OR SUFS OR B1680.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "the complete genome sequence of Escherichia coli K-12";  
 RL Science 277:1453-1474(1997).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-K12 / W3110;  
 RC MEDLINE=97251357; PubMed=9097039;  
 RX Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,  
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,  
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
 RA Nakade S., Nakamura Y., Nishio Y., Oshima T., Saito N.,  
 RA Sempel G., Seki Y., Sivasubraman S., Tagami H., Terada J.,  
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horinouchi T.,  
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
 corresponding to the 28.0-40.1 min region on the linkage map";  
 RL DNA Res. 3:363-377(1996).  
 [3]  
 RN SEQUENCE OF 1-10, CHARACTERIZATION, AND X-RAY CRYSTALLOGRAPHY.  
 RP STRAIN-K12;  
 RC MEDLINE=99262626; PubMed=10329673;  
 RX Mihara H., Maeda M., Fujii T., Kurihara T., Hata Y., Esaki N.;  
 RA "A nifs-like gene, csdb, encodes an Escherichia coli counterpart of  
 RT mammalian selenocysteine lyase. Gene cloning, purification,  
 RT characterization and preliminary x-ray crystallographic studies.";  
 RL J. Biol. Chem. 274:14768-14772(1999).  
 [4]  
 RN GENE NAME.  
 RP STRAIN-K12 / MG1655;  
 RC MEDLINE=99255563; PubMed=10322040;  
 RX Patzer S.I., Hanke K.;  
 RT "Sufs is a Nifs-like protein, and sufd is necessary for stability of  
 RT the zfe-2S FnuF protein in Escherichia coli.";  
 RL J. Bacteriol. 181:3307-3309(1999).  
 [5]  
 RN FUNCTION.  
 RP MEDLINE=20379012; PubMed=10829016;  
 RX Lacourciere G.M., Mihara H., Kurihara T., Esaki N., Stadtman T.C.;  
 RA "Escherichia coli Nifs-like proteins provide selenium in the pathway  
 RT for the biosynthesis of selenophosphate.";  
 RL J. Biol. Chem. 275:23769-23773(2000).  
 [6]  
 RN MOTAGENESIS OF CYS-364.  
 RP MEDLINE=20206733; PubMed=10739946;  
 RX Mihara H., Kurihara T., Yoshimura T., Esaki N.;  
 RA "Kinetic and mutational studies of three Nifs homologs from  
 RT Escherichia coli: mechanistic difference between L-cysteine  
 RT desulfurase and L-selenocysteine lyase reactions.";  
 RL J. Biochem. 127:559-567(2000).  
 [7]  
 RN X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).  
 RP MEDLINE=20150398; PubMed=10684605;  
 RX Fujii T., Maeda M., Mihara H., Kurihara T., Esaki N., Hata Y.;  
 RA "Structure of a Nifs homologue: X-ray structure analysis of CsdB, an  
 RT Escherichia coli counterpart of mammalian selenocysteine lyase.";  
 RL Biochemistry 39:1263-1273(2000)  
 CC -1- FUNCTION: CATALYZES THE RELEASE OF SELENIUM, SO(2), AND SULFUR  
 CC FROM L-SELENOCYSTEINE, L-CYSTEINE SULFINIC ACID, AND SULFUR  
 CC RESPECTIVELY, TO YIELD L-ALANINE; THE REACTIVITY OF THE SUBSTRATES  
 CC DECREASED IN THIS ORDER. FUNCTIONS AS A SELENIUM DELIVER PROTEIN  
 CC IN THE PATHWAY FOR THE BIOSYNTHESIS OF SELENOPHOSPHATE.  
 CC -1- CATALYTIC ACTIVITY: L-selenocysteine + reduced acceptor = hydrogen  
 CC selenide + L-alanine + acceptor.  
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SIMILARITY: BELONGS TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
 CC AMINOTRANSFERASES. CSD SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AE000263; AAC74750.1; -  
 DR EMBL; D90812; BAI15463.1; -  
 DR EMBL; D90811; BAI15457.1; -  
 DR PDB; 1CON; 17-JUL-00.  
 DR PDB; 1TF9; 04-JUL-01.  
 DR Ecogene; EG13962; csdb.  
 DR InterPro; IPR000192; AminoTransf\_class\_V.  
 DR Pfam; PF00266; aminotran\_5; 1.  
 DR PROSITE; PS00595; AA\_TRANSFR\_CLASS\_5; 1.  
 DR Lyase; Pyridoxal phosphate; 3D-structure; Complete proteome.  
 FT BINDING 226 PYRIDOXAL PHOSPHATE.  
 FT ACT SITE 364 364  
 FT MUTAGEN 364 364  
 SQ SEQUENCE 406 AA; 44433 MW; 937AC43C3AD9DBE3 CRC64;  
 C-XA: LOSS OF CYSTEINE DESULFURIZATION.

alignment\_scores:  
 Quality: 101.00 Length: 362  
 Ratio: 0.591 Gaps: 17  
 Percent Similarity: 47.238 Percent Identity: 23.204

## Alignment\_block:

US-09-303-518D-125 x CSDB\_ECOLI ..  
 Align seg 1/1 to: CSDB\_ECOLI from: 1 to: 406

```

229 GCCGGATTGACCGTGC.....GAAA 251
52 AlaIaValInIArgGlyLeuSerAlaGlnAlaThrGly 68
252 G.....CGGTAATTCAGTCGATGCGGTGAAAGCA 289
68 SMeGlnAsnValArgLysArgAlaSerLeuPheIleAsnAlaArgSer 85
290 ACGAGAAATCGAGTTGAACGCTACGCACTGAGCGGTG..... 330
85 IagIuGluLeuValPheValArgGlyThrThrIuGluIleAsnLeuVal 101
331 GCAAACTTAAGCGCGAAGAAGTGGCGG.....AAGCTGATCCA 371
102 AlaAsnSerTrpLysSerAsnValArgAlaGlyAspAsnIleIle 118
372 ATCCGGTTGTGACGCTGCGCGACCGCGTCCAGCAAAATTCG 421
118 eSerGlnMetGlnInIshAlaAsnIleValProTrpGlnMetLeuCys 135
422 CC...GTGATGCCGAGCGCGTGGCATGTCGATGCGATGAGCA 468
135 IaaArgValGlyAlaGlyLeuArgValIleProLeuAsnProAspGlyThr 151
469 AATCCGCTGCTGCGACCGCTACGGTCATTAACAAGAAGCGCGAG 518
152 LeuGlnLeuGluThrLeuProThrLeuPhe.....As 162
519 TTTCAAACCGCGCGCTGTGATGAGCGTTGACCGAA..... 558
162 pGluLysThrArgLeuLeuAlaIleThrHisValSerAsnValLeuGly 179
559 .....CGCAAAATCCATGTTTGAAGCGAGCTGGCGCA 591
179 hGlnAsnProLeuAlaGlnMetIleThrLeuAlaInIshGlnIshGlyAla 195
592 GACGTGCGCGTGAATAATGCTGCGCAACATCGAAACACATGAA..... 633
196 LysValLeuValAspGlyAlaGlnAlaValMetHisIshProValAsp 212
634 .....TTGGCGCGCGCGCATCTGCGG 655
212 IGlAlaLeuAspCysAspPheTyrValPheSerGlyHisLysLeuTyrG 229
656 GTTTGAGTGGCAGCGCACATTCATTCATGAGCGCGT..... 693
229 LyrThrGlyLeuGlyIleLeuTyrValLysGlnAlaLeuGlnGlu 245
694 .....GGCGGAATTAACCGTGGACCATCAATTATCA 728
246 MetProProTrpGluGlyGlySerMetIleAlaThrValSerLeu 262
729 ACATGTAATTACCATTTGGCGGTG.....TTTCAACAGCGCGT 769
262 rGluGlyThrThrTrpThrLysAlaProTrpArgPheGlnAlaGlyThr 279
770 TGAACACCGAGCGCGTGTGCGGTGCGTGTCTCAAGTCAACAACG 819
279 rGlnThrGlyGlyIleIleLeuGlyAla..AlaLeuGluTyrValS 295
820 CGCTCTTGCGCGTACCGGTTTGGGTGCGAAGTATCGAAATTAAGT 869
295 eAlaLeu.....GlyLeuAsnIleAlaGlyTyr..... 305
870 GCAATGTGTGACACAGAACCGCGTGAATTCGCGTTCGATTTGACG 919
306 .....GluGlnAsnLeuMetHisTyrAlaLeuSerG 316
920 GCGCGATTACACAGCGCGCGCATTTATTTGGAGCGTACCAATTCAG 969
316 nLeuGlnSerValProAspLeuThrLeuTyrGly.....ProGlnAsn 330
970 ATTTCCGTTATCGAAGAAGCGCGACGAAGACGTGTCGCGGTGTCG 1019
331 .....ArgLeuGlyVal 334
1020 GCCGACCGCGAACAATATCTCAATGACGCGTACACCGCGCATTTCC 1069
335 ILeAlaPheAsnLeuGlyLysHisIshAlaTyrAsp.ValGlySerPhe 351
1070 TGAATAACAACCTTCAAGTTCACACAGCGCGTCAAGCGCGCGAGCG 1119
351 euAspAsn.....TyrGlyTAlaValArgThrIshHis 363
1120 GCCATGTCGCGATTTGGTACTTACGAGCGGTG 1152
364 CysAlaMetProLeuMetAlaTyrTyrAsnVal 374

```

THE UNIVERSITY OF CHICAGO

1964